2012-05-16

Applications of Lexical Link Analysis Web Service for Large-Scale Automation, Validation, Discovery, Visualization, and Real-Time Program Awareness

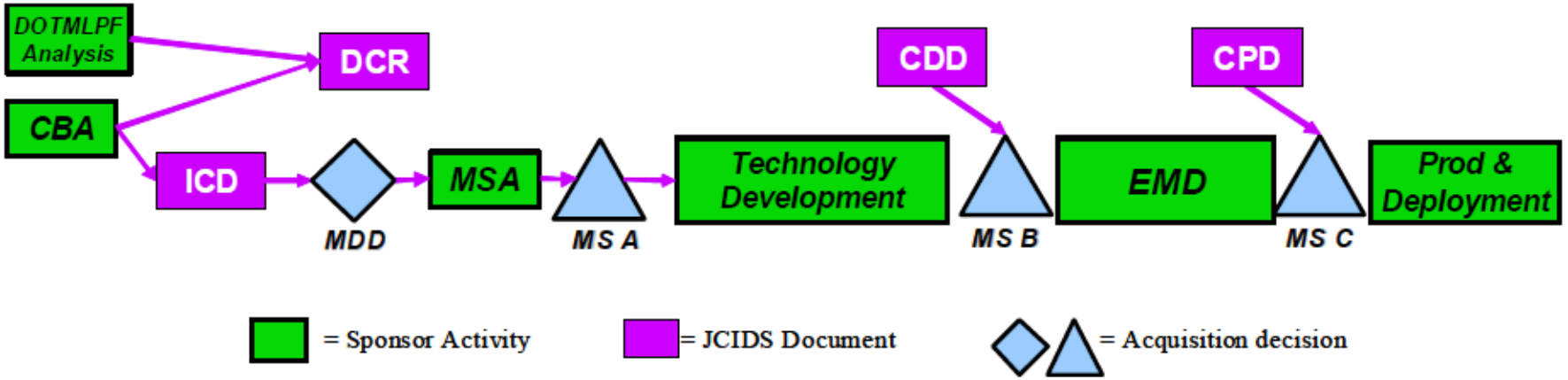
Acquisition Research Program:
Creating Synergy for Informed Change

APPLICATIONS OF LEXICAL LINK ANALYSIS WEB SERVICE FOR LARGE-SCALE AUTOMATION, VALIDATION, DISCOVERY, VISUALIZATION AND REAL-TIME PROGRAM-AWARENESS

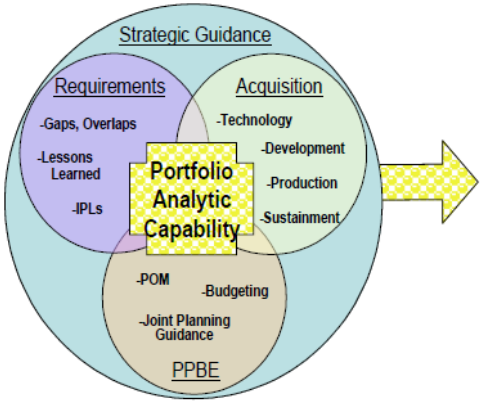
May 16-17, 2012

Dr. Ying Zhao, Dr. Douglas J. MacKinnon, Dr. Shelley P. Gallup
Research Associate Professors
Distributed Information Systems Experimentation, Naval Postgraduate School

Critical Needs: Automation, Validation and Discovery



JCIDS Process and Acquisition Decisions
(J-8 CJCSI 3170.01G)(JCIDS, 2009)



Multiple Portfolio Views:

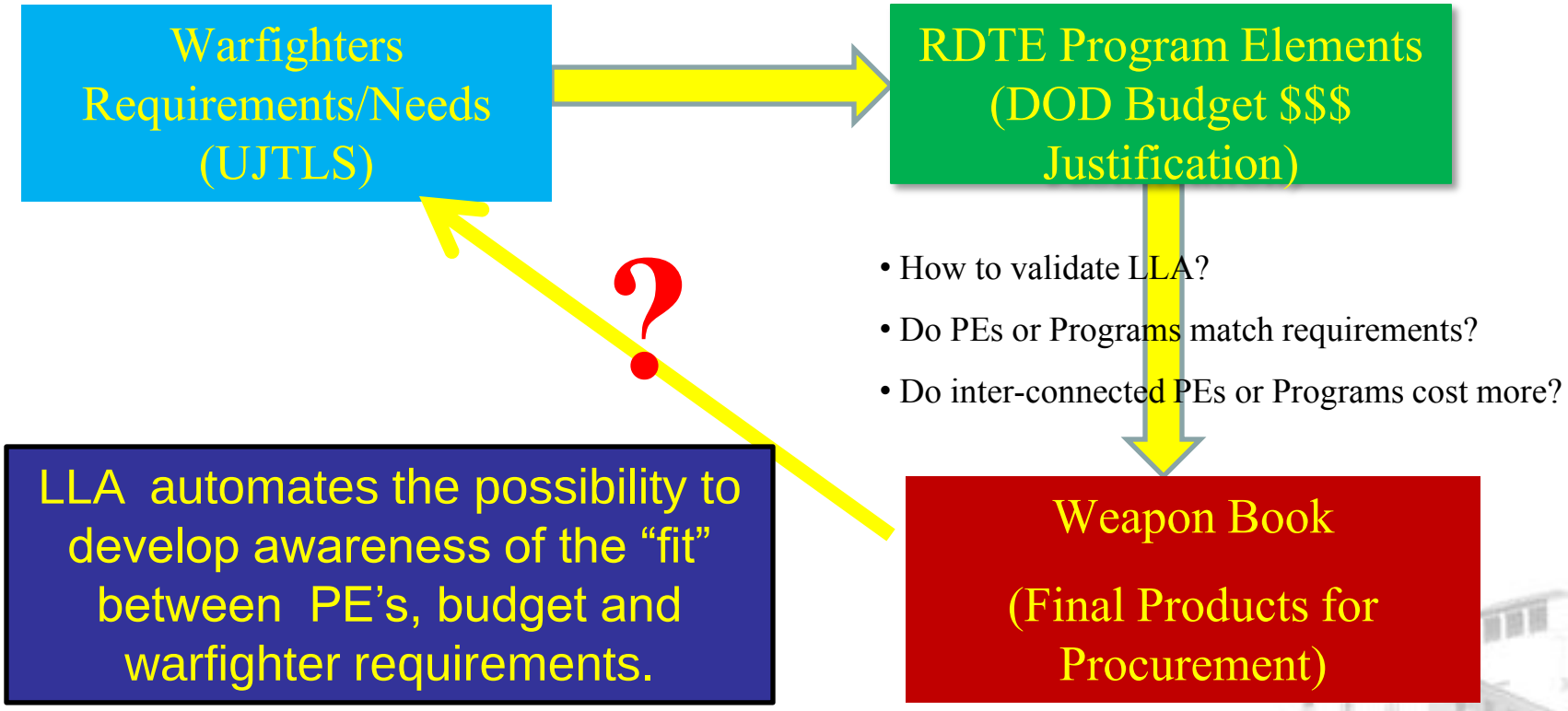
- Systems vs. Capabilities
- Investment vs. Capabilities
- System Context
- Highly dependent programs (Joint Enablers)
- Procurement Optimization
- S&T vs. future needs
- Sustainment Efficiency
- Market Value

- Data are too voluminous, unformatted and unstructured!
- Need to leverage automation
  - Extract relations among PE, MDAP, and ACATII
  - Extract costs

Research Question

How can the information that emerges from the acquisition process be used to produce overall *awareness* of the *fit* between programs/projects/systems and verify *needs* for which they were intended?

LLA Methodology Can Help!



METHODS

System Self-Awareness (SSA)

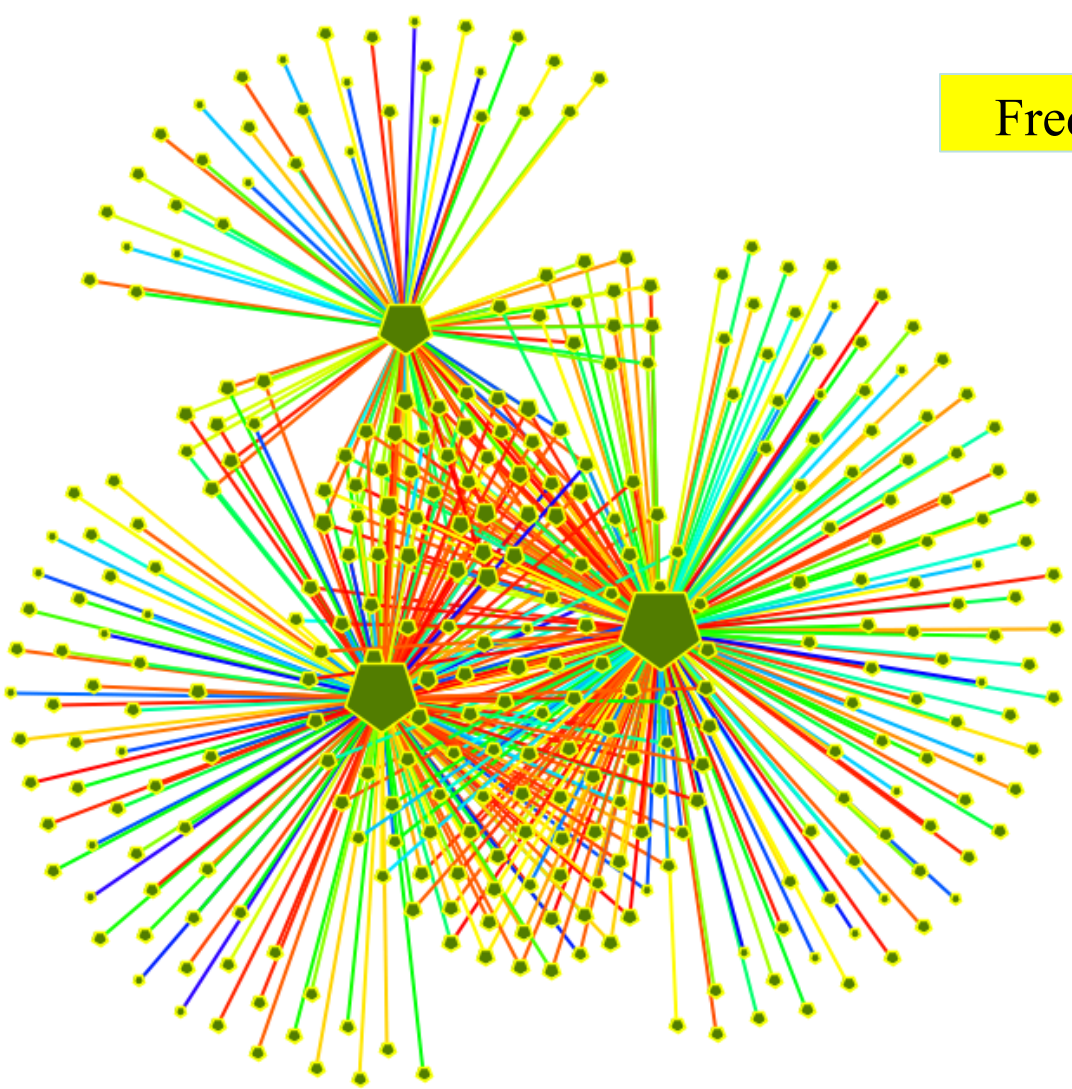
- *Awareness*
  - The cognitive interface between decision makers and a complex system, expressed in a range of terms or “features,” or specific vocabulary or “lexicon,” to describe the attributes and surrounding environment of the system.
- System Self-awareness
  - Complex system’s ability to assess itself within a global context
  - Examples
    - Authority
    - Expertise

Text Analysis

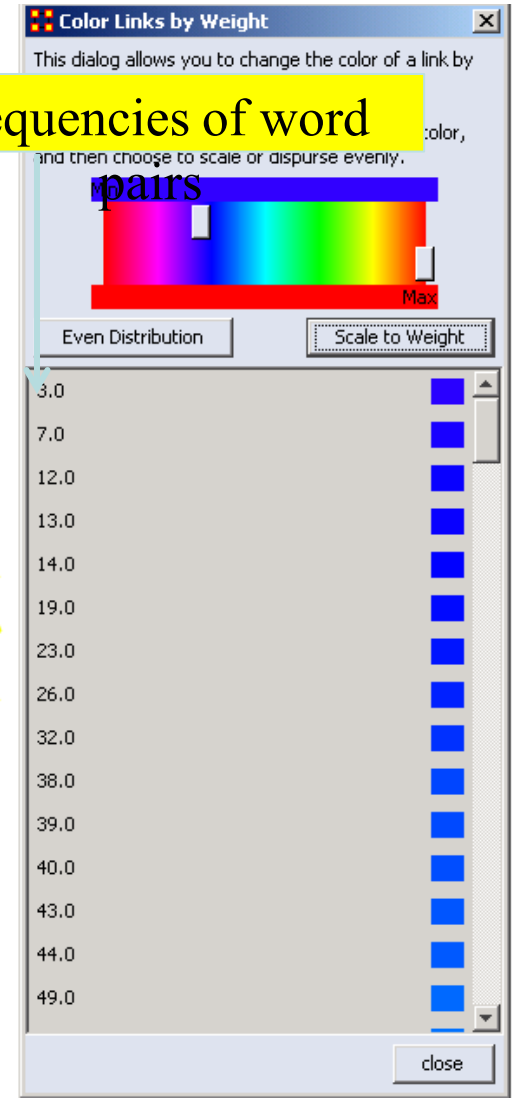
There are three methods

- Linguistics based methods
  - InXight
- Statistical co-occurrence
- Representation
  - Bag-of-Words (BOW)
  - Text-as-Network (TAN)

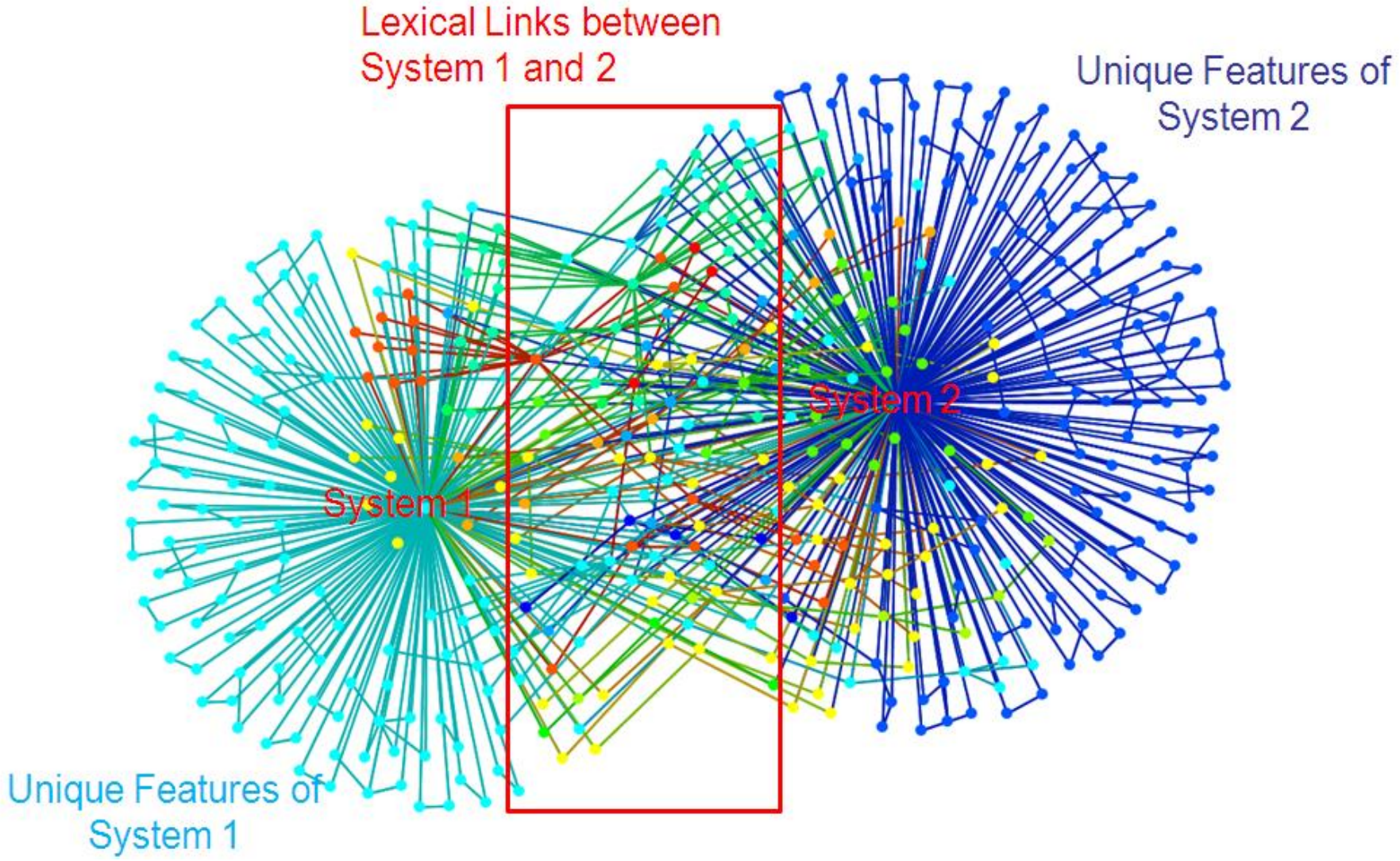
LLA: Bi-gram co-occurrence word pair networks



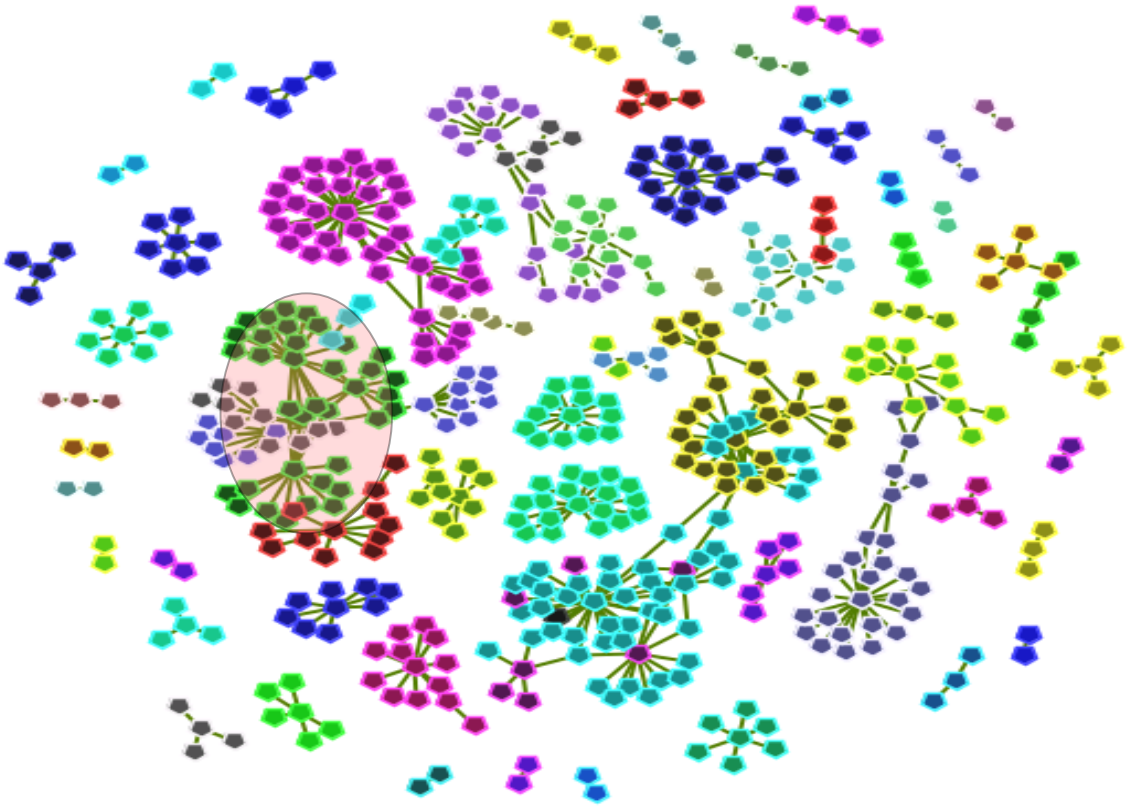
Frequencies of word pairs



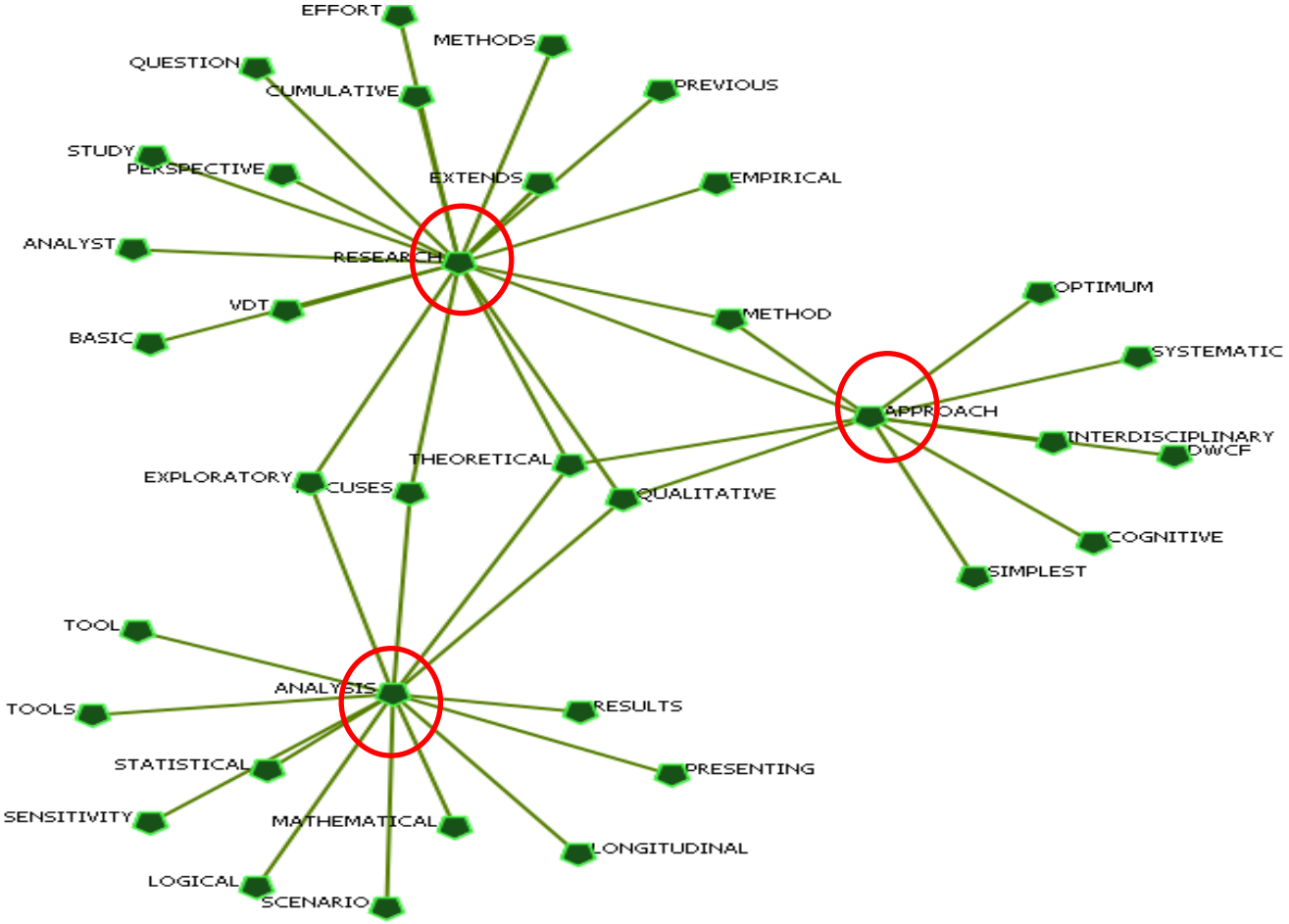
Comparing Two Systems using LLA



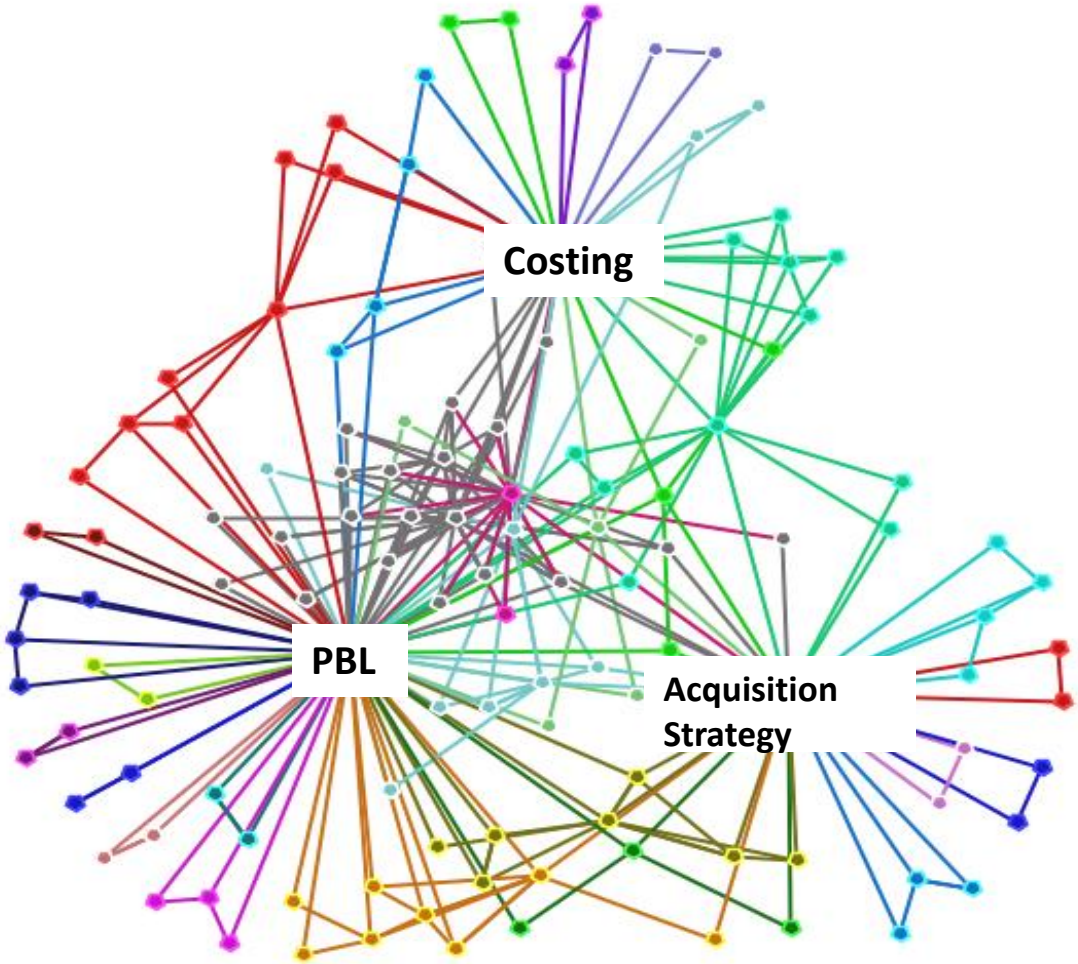
Themes



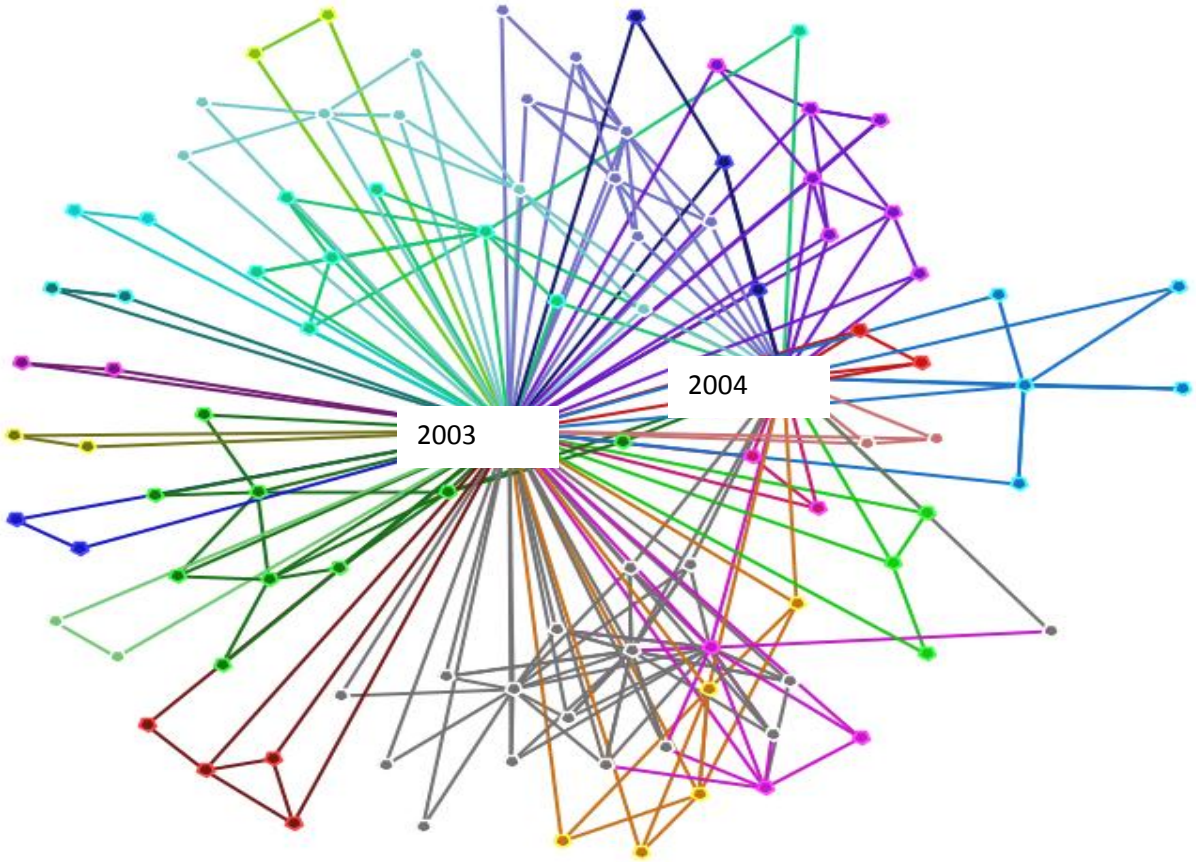
Details


EFFORT
METHODS
QUESTION
CUMULATIVE
PREVIOUS
STUDY
PERSPECTIVE
EXTENDS
EMPIRICAL
ANALYST
RESEARCH
VDT
METHOD
BASIC
OPTIMUM
SYSTEMATIC
APPROACH
INTERDISCIPLINARY
DWCF
THEORETICAL
EXPLORATORY
CUSES
QUALITATIVE
COGNITIVE
SIMPLEST
TOOL
ANALYSIS
TOOLS
RESULTS
STATISTICAL
PRESENTING
SENSITIVITY
MATHEMATICAL
LONGITUDINAL
LOGICAL
SCENARIO

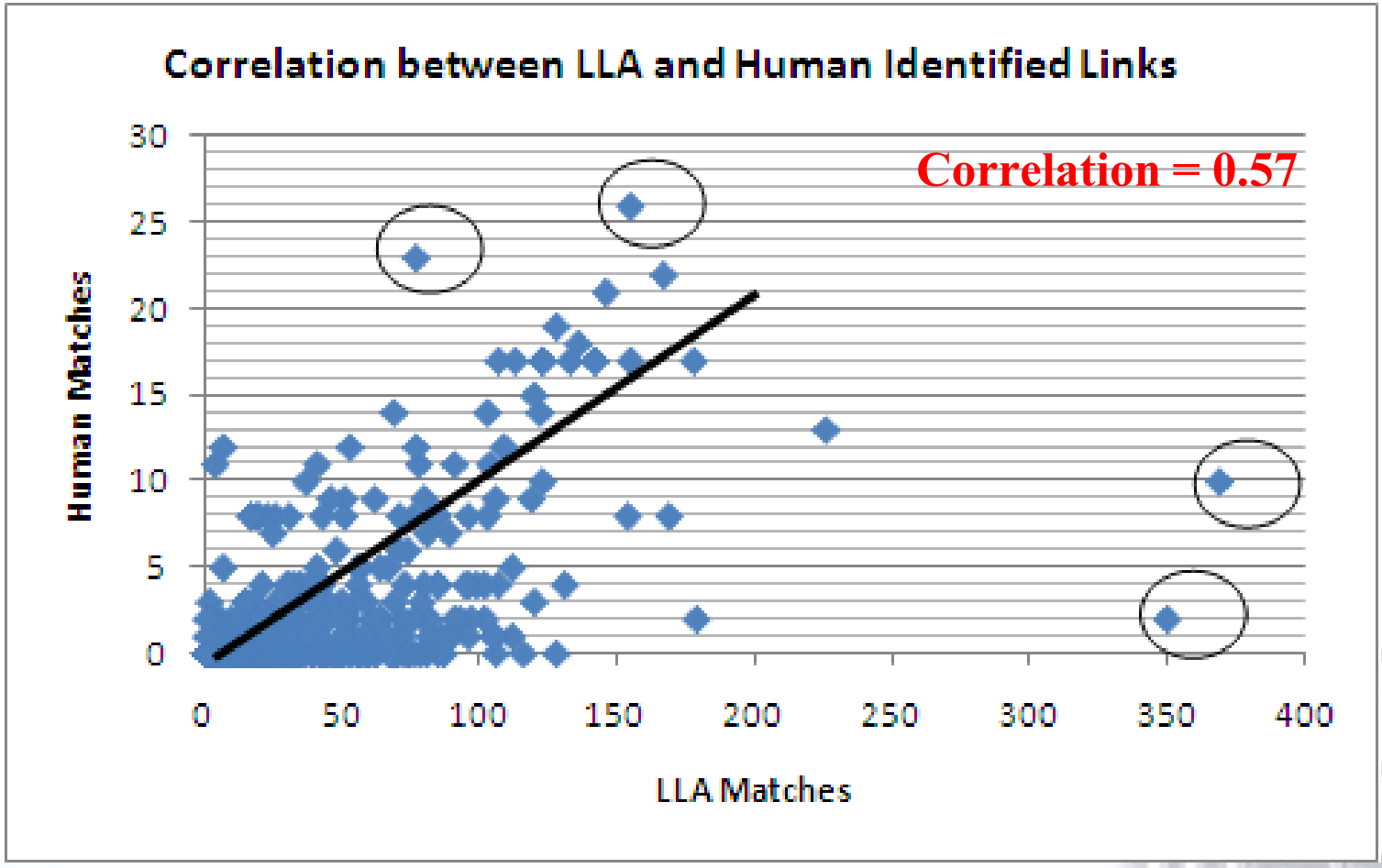
Comparing Categories



Compare Time Points



Phase I Results: Validation of LLA

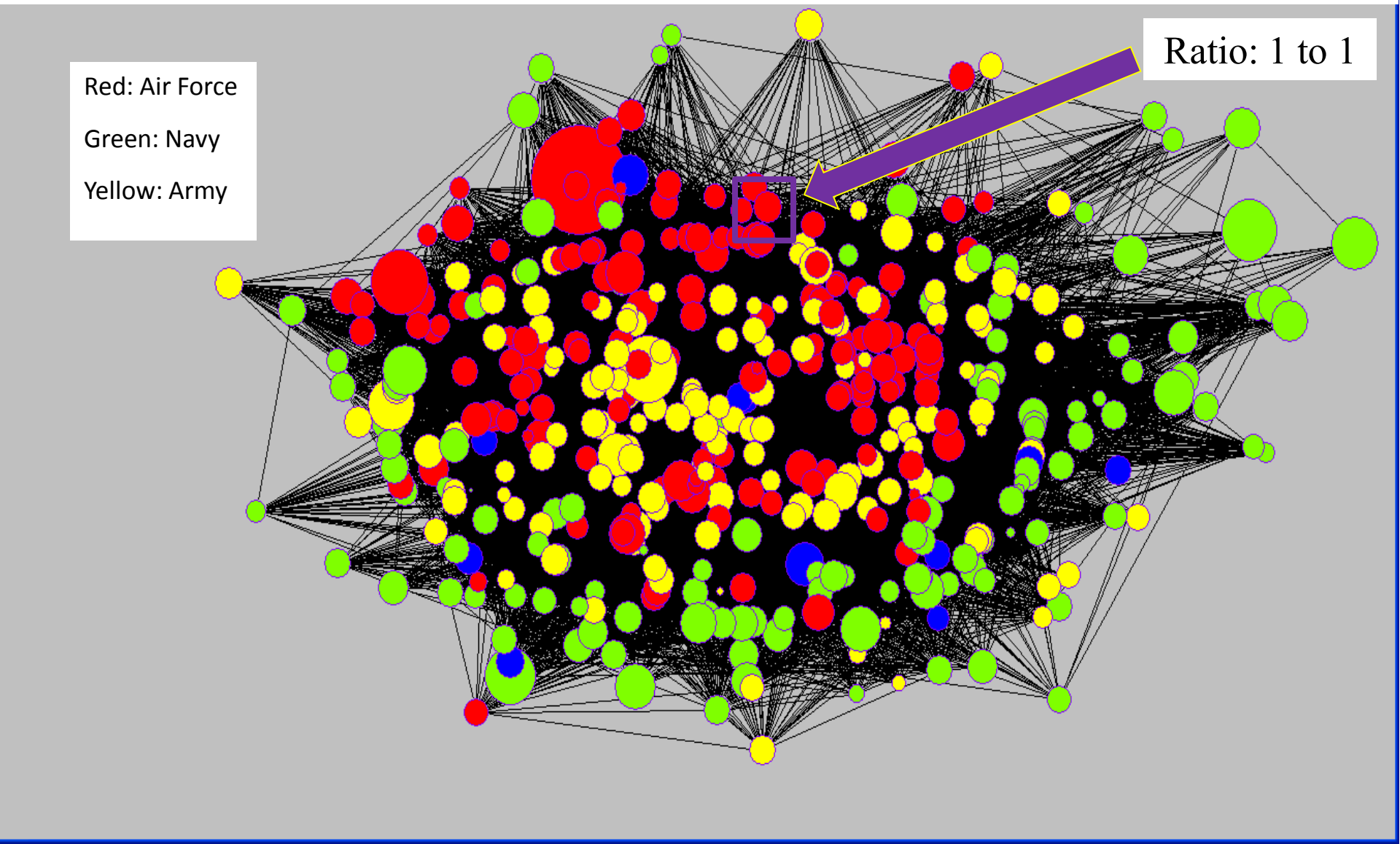


LLA Benefits

- High correlation exists between LLA results and human analyses
  - Establishes the potential to use lexical links to rank documents, concepts and themes.
- LLA can also focus on ***innovations and uniqueness*** of the analyzed documents
  - Other ranking techniques which typically sort documents based on the *popularity* or *authority, are* not based on semantics
    - E.g. PageRank by Google

Trend Analysis

Semantic Network: Size of Nodes = 2009 Cost / 2008 Cost



Phase III Objectives

- Build at least two use cases of applications of Lexical Link Analysis Web Service for large-scale automation, validation, discovery, visualization, and real-time program awareness.
- Demonstrate the methodology for assisting the DoD-wide effort of integrating and maintaining authoritative and accurate acquisition data services in both legacy and new platforms.

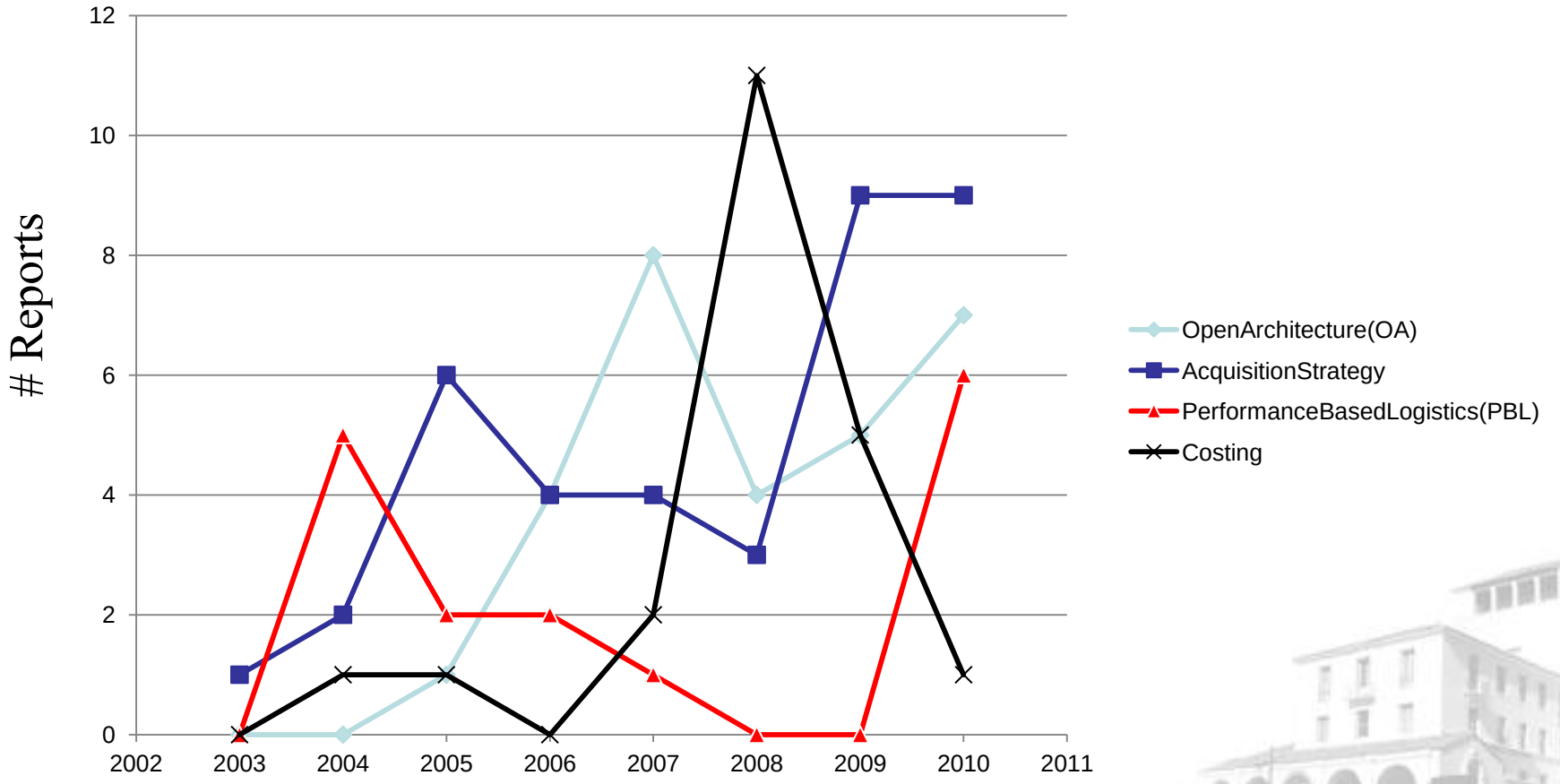
Acquisition Research Program

- 740 publications (from 2003 to 2010) from the website http://www.acquisitionresearch.net
- Pre-defined categories
  - "There are ~160 categories, e.g. Acquisition Strategy, Costing, Open architecture, Systems of Systems

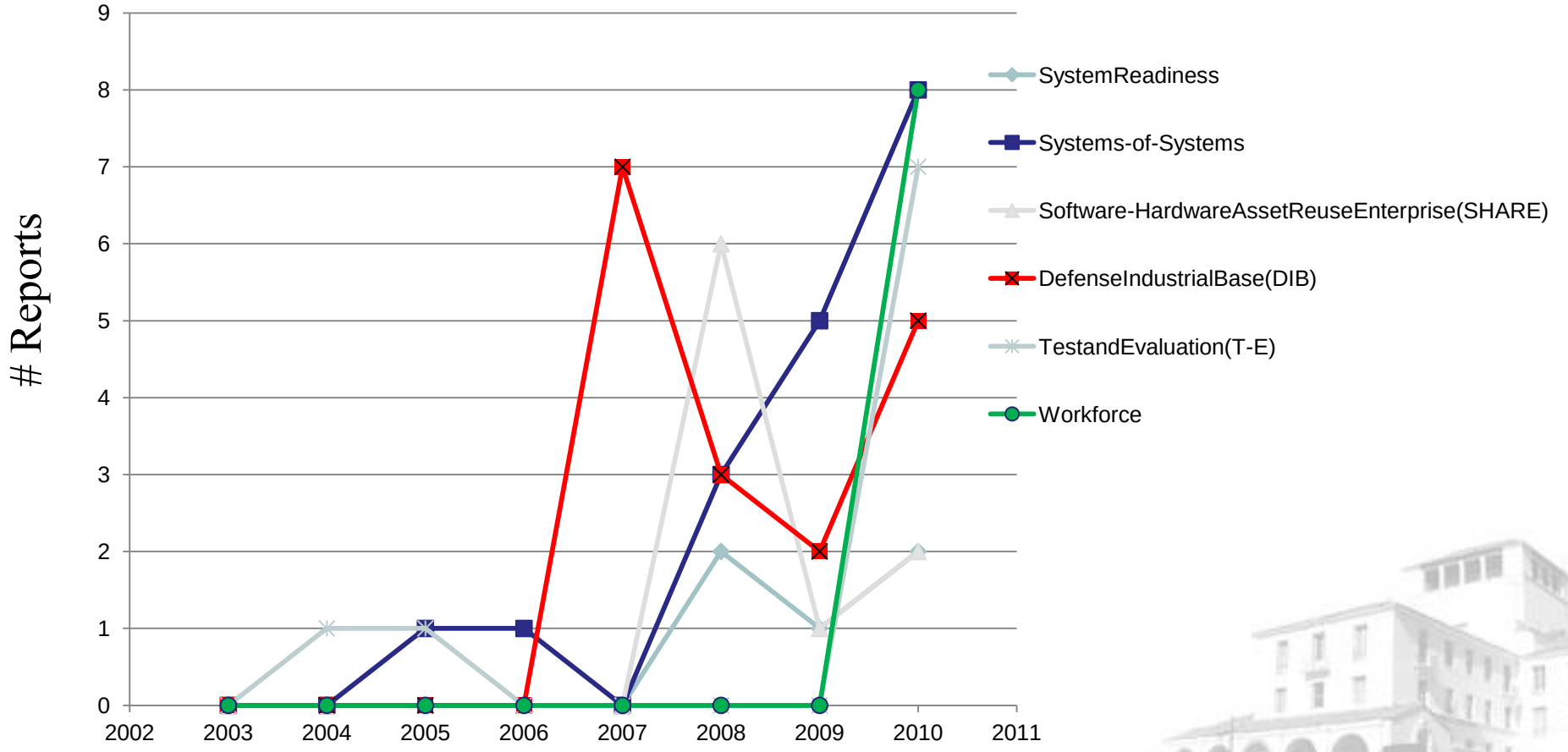
Year	# of Reports	# of Categories
2003	8	6
2004	27	17
2005	61	34
2006	62	29
2007	143	63
2008	144	68
2009	127	61
2010	184	65

ARP Reports from 2003 to 2010

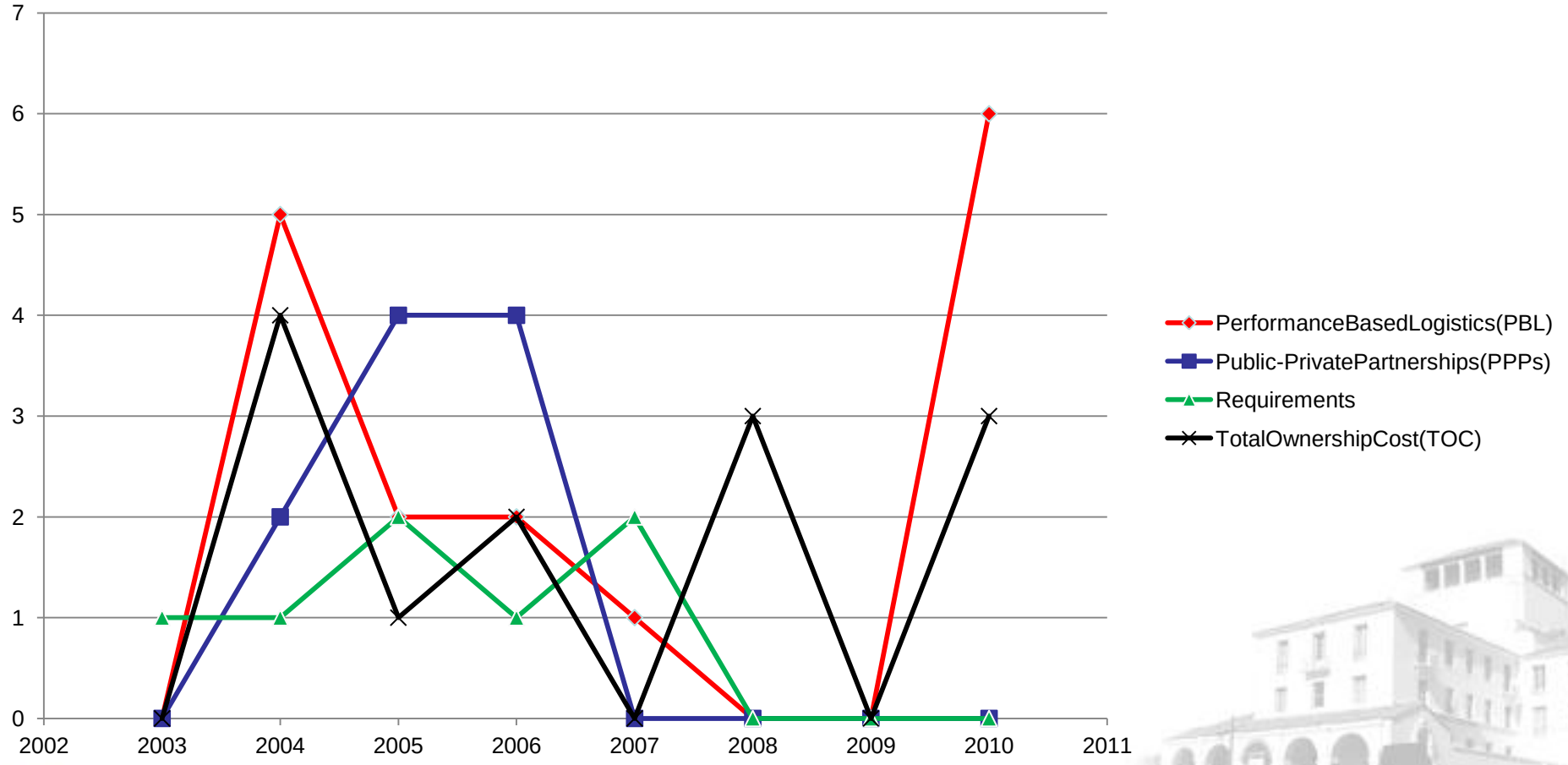
Steady Categories



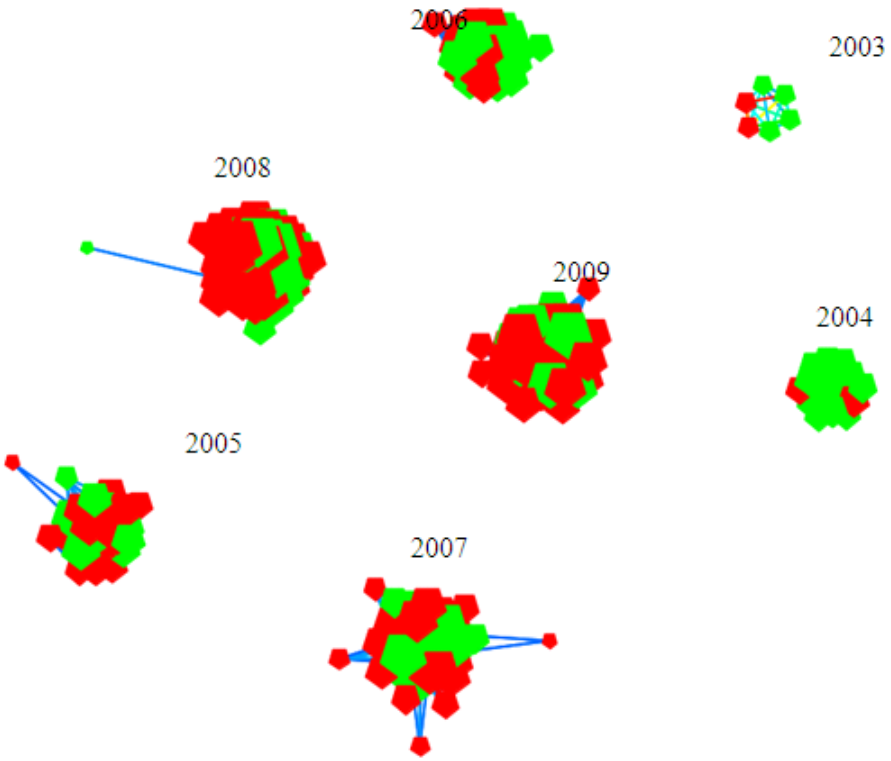
New and Emerging Categories



“Sunset” Categories

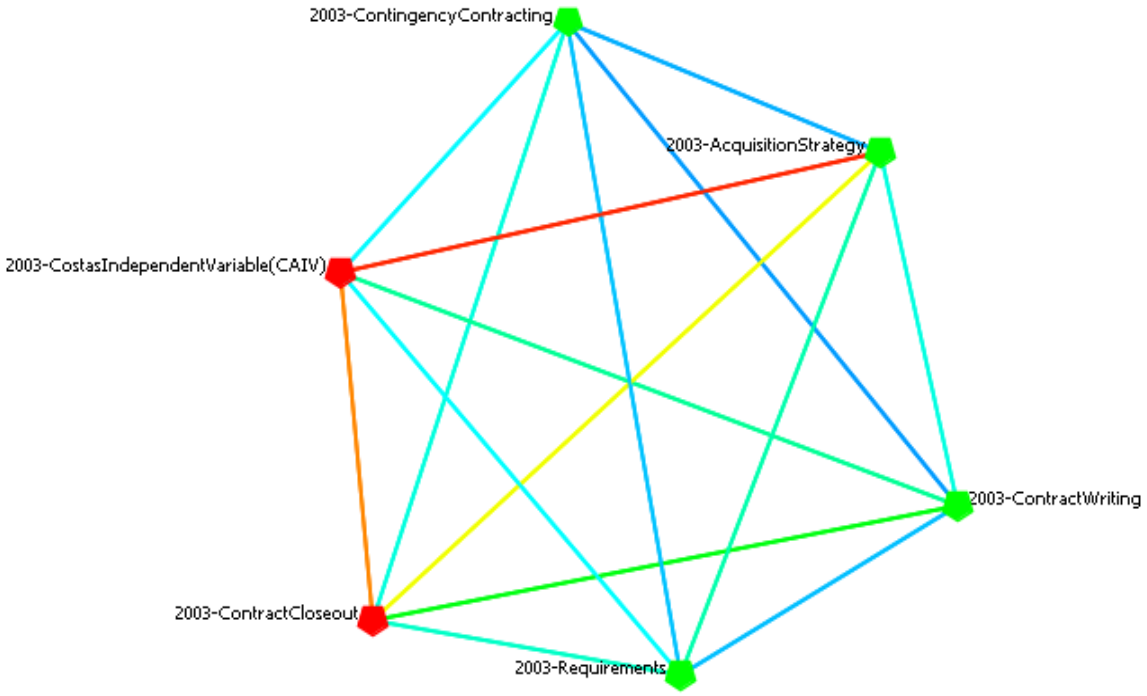


Details



- 240 objects (combinations), e.g. 2003-AcquisitionStrategy and 2004-Outsourcing,.
- For each combination
  - Label 1 (*kept*), if the associated category was continued in the following year, e.g. 2003-AcquisitionStrategy are both 2004-AcquisitionStrategy is also one
  - Label 0(*deleted*), if the associated category was not continued in the next year, e.g. 2003-ContractCloseout is an existing category, but 2004-ContractCloseout is not -- no reports were classified in the ContractCloseout category in 2004
- Semantic networks for each year
  - Green – 1(kept)
  - Red – 0 (deleted)

2003



Increased (growth, green)

- Acquisition Strategy
- Contract Writing
- Requirements
- Contingency Contracting

Reduced

Decreased (red)

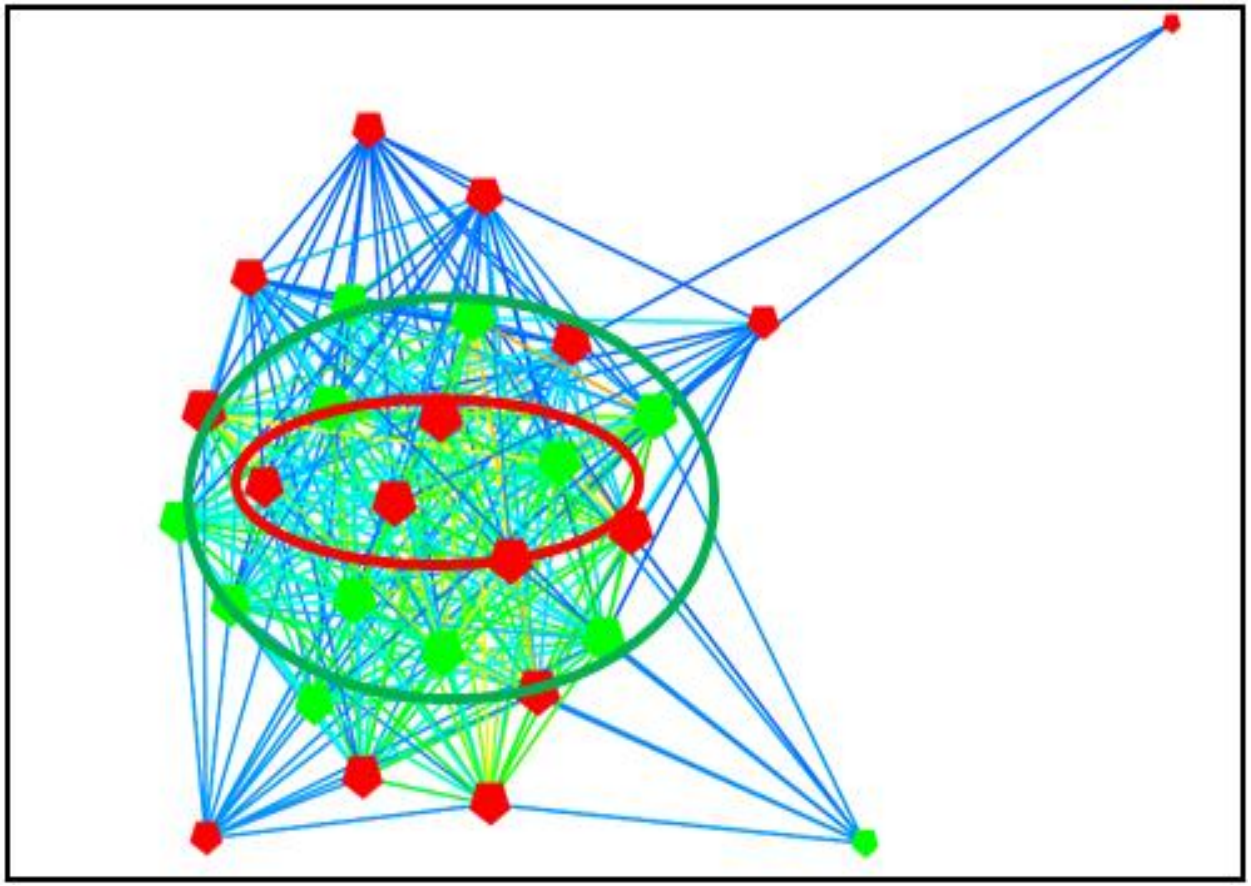
- Cost Independent Variable
- Contract Closeout

Statistical Significant Tests

	Total	Deleted	Kept	Kept/Total	
Group A (LLA Score<7)	76	53	23	0.30	
Group B (LLA Score>=7)	169	84	85	0.50	p=0.0017
Group C (Top Ranked in Total Degree)	76	47	29	0.38	
Group D Rest	169	90	79	0.47	p=0.1053

- **Green nodes have stronger (LLA scores higher) but fewer links (Total degrees lower)**

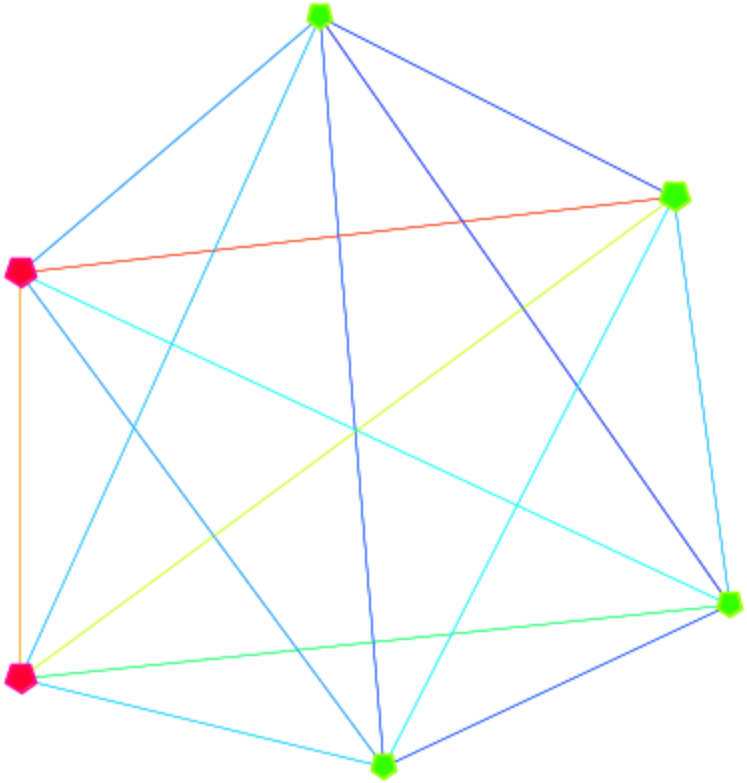
Ring of Emergence



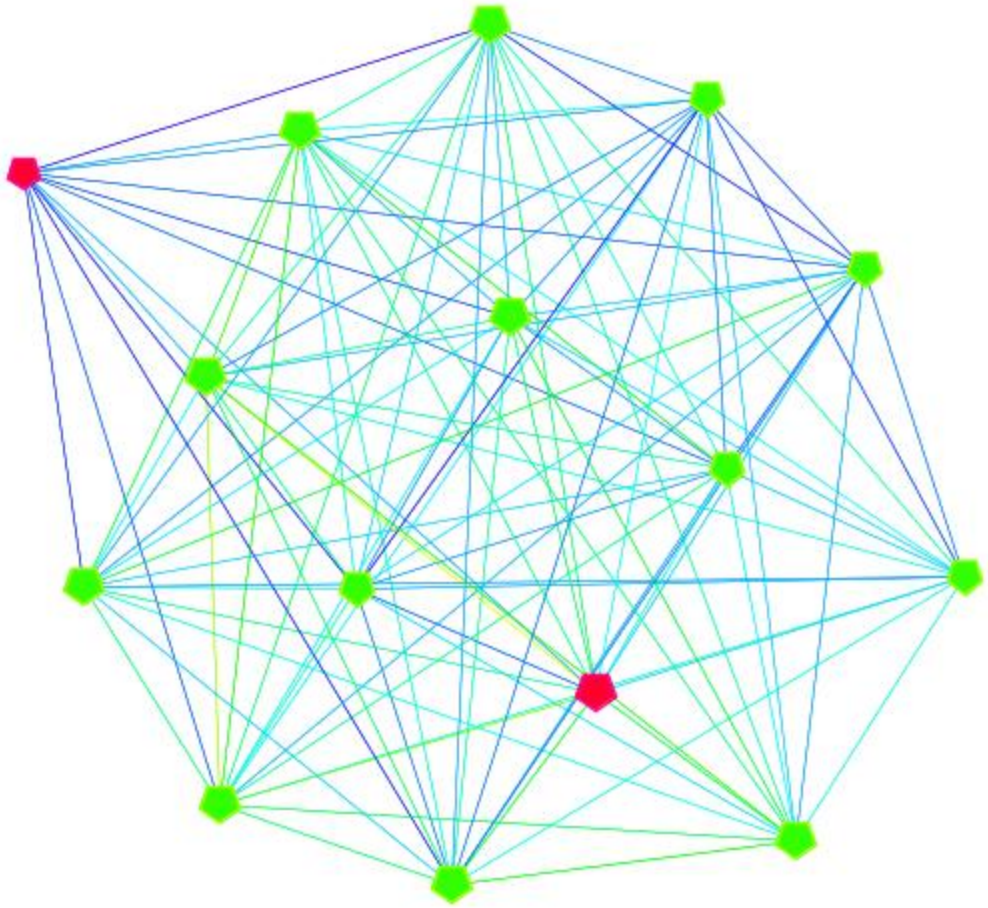
Green nodes have stronger (LLA scores higher) but fewer links (Total degrees lower)

- Green nodes not in the centers but in a ring
- Associate with hotter nodes (less blue)

2003

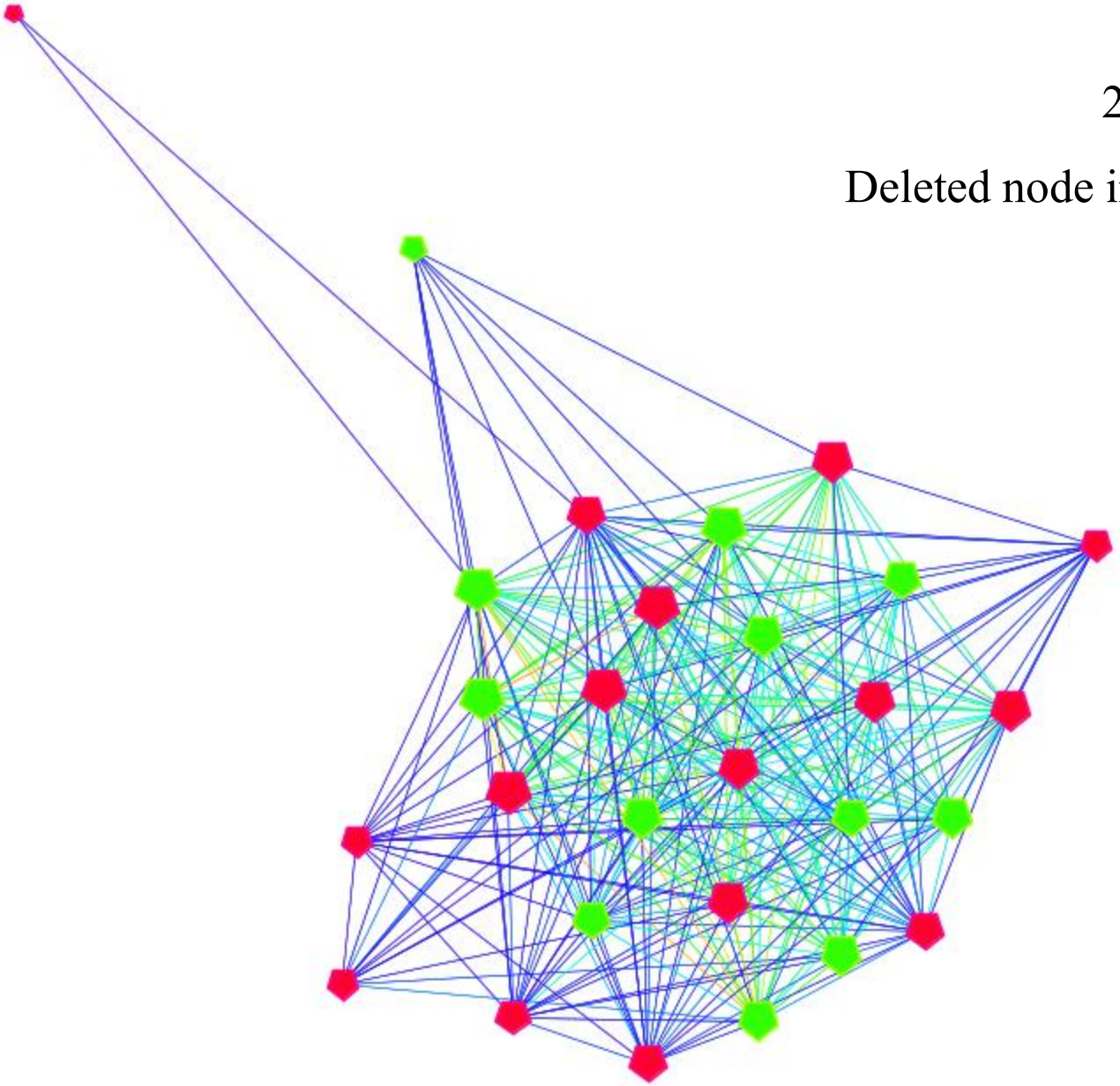


2004



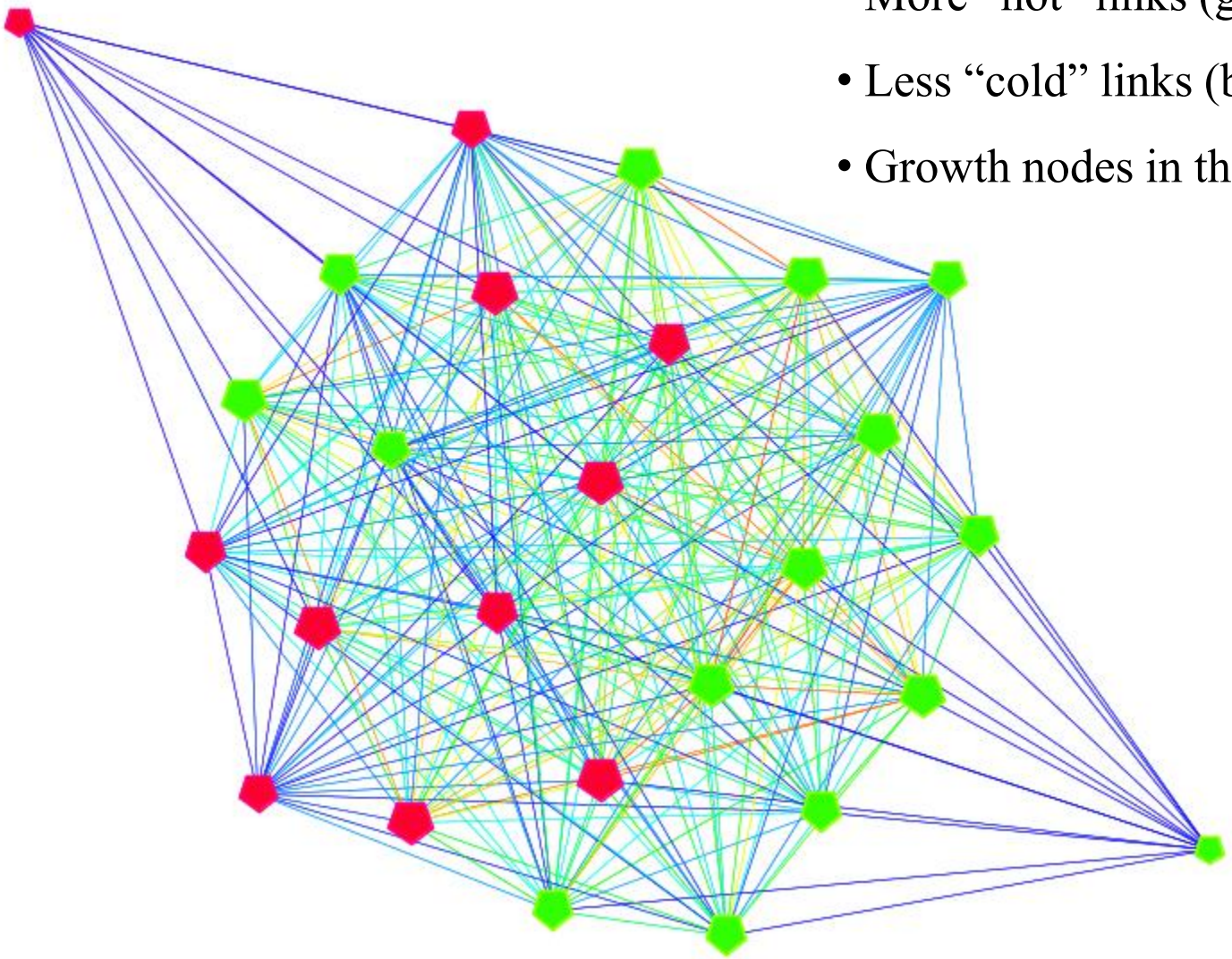
2005

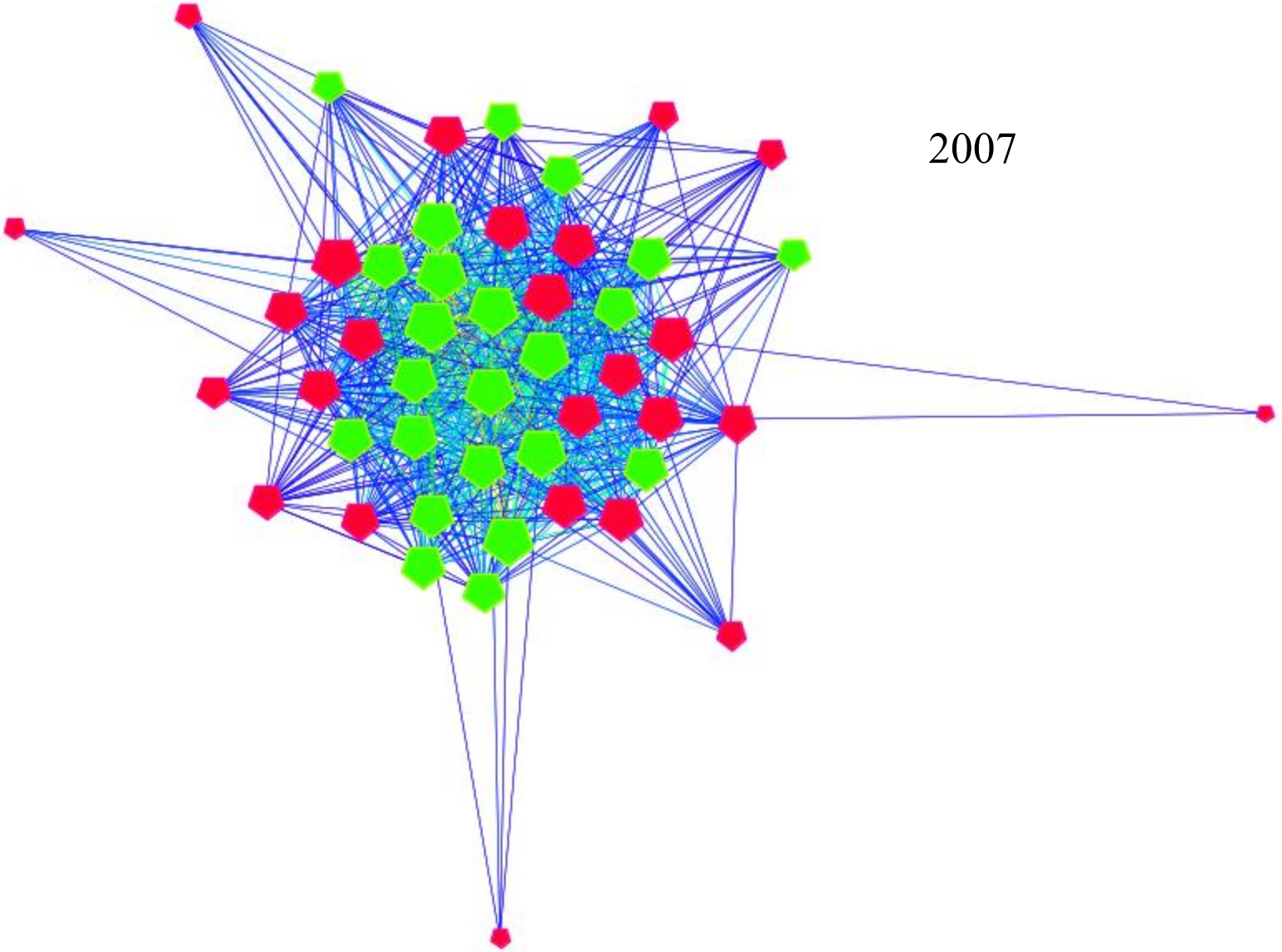
Deleted node in the “cold” areas



2006: More kept nodes (red) than deleted

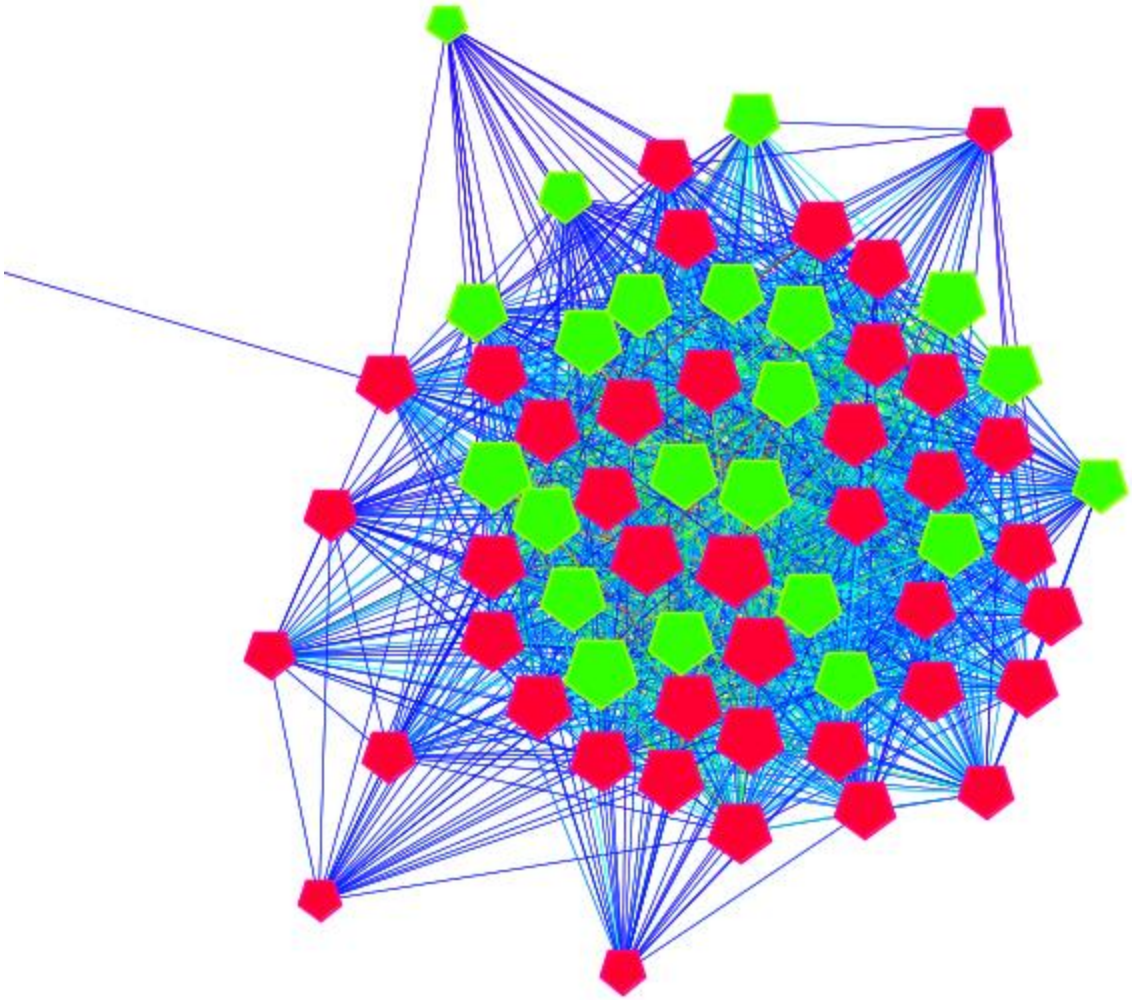
- More “hot” links (green and red)
- Less “cold” links (blue)
- Growth nodes in the “hot link” areas



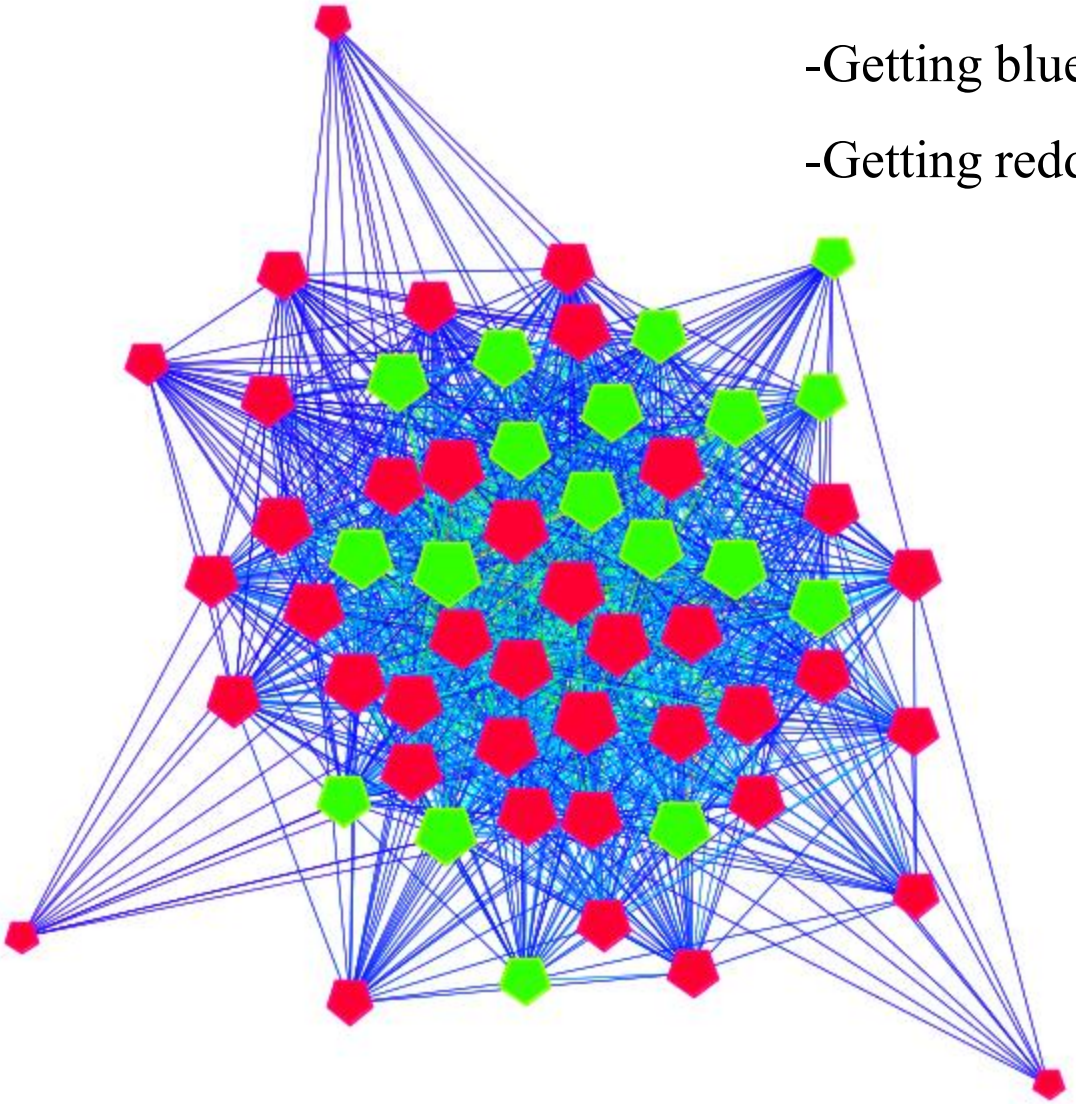


2007

2008



2009



-Getting bluer: smaller LLA scores

-Getting redder: more deleted nodes

Future Work and Why It is Important

- Is the DoD ARP system Pareto efficient?
  - How to use LLA and Collaborative Learning Agents (CLA) to make decisions that achieve an overall more efficient system
    - E.g. a DOD acquisition search system that can reinforce the diversity, uniqueness, and innovations of the technologies and investments, not just based on authorities, popularities. This could lead to a more Pareto efficient or *swarm intelligent* selection of acquisition programs

Seeking to Work with ARP Partners

- Accurate and authoritative data services in both legacy and new platforms into strategic decision-making knowledge

1. PEs: http://www.dtic.mil/descriptivesum/
2. MDAPs & ACATIIs: http://comptroller.defense.gov/defbudget/fy2008/fy2008_weabook.pdf
http://www.fas.org/man/dod-101/sys/land/wsh2007/index.html
http://www.acq.osd.mil/ara/am/sar/
3. UJTLs: http://www.dtic.mil/doctrine/jel/cjcsd/cjcsm/m350004d.pdf

- According to the Enterprise Information & OSD Studies, Office of the Under Secretary of Defense - Acquisition, Technology & Logistics (OUSD AT&L), these data sources provide the DoD-wide acquisition community with authoritative and accurate data services among others such as DAMIR(http://www.acq.osd.mil/damir/), ARA(http://www.acq.osd.mil/ara, and Selected Acquisition Report (SAR) (http://www.acq.osd.mil/ara/am/sar).

Acquisition Research Program:
Creating Synergy for Informed Change

APPLICATIONS OF LEXICAL LINK ANALYSIS WEB SERVICE FOR LARGE-SCALE AUTOMATION, VALIDATION, DISCOVERY, VISUALIZATION AND REAL-TIME PROGRAM-AWARENESS

May 16-17, 2012

Dr. Ying Zhao, Dr. Douglas J. MacKinnon, Dr. Shelley P. Gallup
Research Associate Professors
Distributed Information Systems Experimentation, Naval Postgraduate School

BACK-UP SLIDES

Statistical Test Example: QAP Correlation
Quadratic Assignment Procedure [QAP; Hubert & Schultz, 1976]

QAP Correlations

		1 lla_n	2 lla_n	3 lla_n	4 lla_n	5 lla_n	6 lla_n	7 lla_n	8 lla_n
1	lla_network_1_2010-AcquisitionStrategy	1.000	0.174	0.156	0.155	0.036	0.111	0.020	0.062
2	lla_network_1_2003-AcquisitionStrategy	0.174	1.000	0.447	0.149	0.052	0.119	0.043	0.089
3	lla_network_1_2004-AcquisitionStrategy	0.156	0.447	1.000	0.111	0.047	0.119	0.051	0.080
4	lla_network_1_2005-AcquisitionStrategy	0.155	0.149	0.111	1.000	0.156	0.084	0.034	0.088
5	lla_network_1_2006-AcquisitionStrategy	0.036	0.052	0.047	0.156	1.000	0.067	0.036	0.056
6	lla_network_1_2007-AcquisitionStrategy	0.111	0.119	0.119	0.084	0.067	1.000	0.097	0.123
7	lla_network_1_2008-AcquisitionStrategy	0.020	0.043	0.051	0.034	0.036	0.097	1.000	0.286
8	lla_network_1_2009-AcquisitionStrategy	0.062	0.089	0.080	0.088	0.056	0.123	0.286	1.000

QAP P-Values

		1 lla_n	2 lla_n	3 lla_n	4 lla_n	5 lla_n	6 lla_n	7 lla_n	8 lla_n
1	lla_network_1_2010-AcquisitionStrategy	0.000	0.020	0.020	0.020	0.020	0.020	0.020	0.020
2	lla_network_1_2003-AcquisitionStrategy	0.020	0.000	0.020	0.020	0.020	0.020	0.020	0.020
3	lla_network_1_2004-AcquisitionStrategy	0.020	0.020	0.000	0.020	0.020	0.020	0.020	0.020
4	lla_network_1_2005-AcquisitionStrategy	0.020	0.020	0.020	0.000	0.020	0.020	0.020	0.020
5	lla_network_1_2006-AcquisitionStrategy	0.020	0.020	0.020	0.020	0.000	0.020	0.020	0.020
6	lla_network_1_2007-AcquisitionStrategy	0.020	0.020	0.020	0.020	0.020	0.000	0.020	0.020
7	lla_network_1_2008-AcquisitionStrategy	0.020	0.020	0.020	0.020	0.020	0.020	0.000	0.020
8	lla_network_1_2009-AcquisitionStrategy	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.000

QAP statistics saved as datafile QAP Correlation Results

Program Elements: Center of Many Things

http://www.dtic.mil/descriptivesum/

Exhibit R-2a, RDT&E Project Justifica

DATE: May 2009

BUDGET ACTIVITY: 05 System Development and Demonstration (SDD)

PE NUMBER AND TITLE: 0604602F Armament/Ordnance Development

PROJECT NUMBER AND TITLE: 5361 Stores-Aircraft Interface

Cost ($ in Millions)	FY 2008 Actual	FY 2009 Estimate	FY 2010 Estimate	FY 2011 Estimate	FY 2012 Estimate	FY 2013 Estimate	FY 2014 Estimate	FY 2015 Estimate	Cost to Complete	Total
5361 Stores-Aircraft Interface	0.000	0.000	6.685	[illegible]	0.000	0.000	0.000	0.000	Continuing	TBD
Quantity of RDT&E Articles	0	0	0	0	0	0	0	0		

In FY 2010, Project 5361, Stores-Aircraft Interface (new), efforts were transferred from PE 0605011F, RDT&E for Aging Aircraft, Project 654685, Universal Armament Interface (UAI), in order to properly fund the maturing technology.

(U) A. Mission Description and Budget Item Justification

Universal Armament Interface (UAI) is an Air Force initiative to develop, enhance, and implement standardized interfaces in aircraft, weapons and mission planning to support integration of weapons independent of aircraft Operation Flight Program (OFP) cycles. UAI is currently being implemented on the F-15E and F-16 Block 40/50 aircraft, Small Diameter Bomb (SDB) I and II, Joint Direct Attack Munition (JDAM), Joint Air-to-Surface Stand-off Missile (JASSM) and Precision Guided Munitions Planning Software (PGMPS). Additional aircraft and weapons have program plans to implement UAI. The UAI program office is responsible for development and enhancement of the standard, provision of certification tools (test assets) and implementation support to aircraft and weapons.

The UAI efforts were transferred (1) to ensure continued funding for UAI through the FYDP (PE 0605011F will be zeroed out in FY 2010 due to higher Air Force priorities), and (2) to properly fund the maturing technology. The new project number is established to provide greater visibility into UAI's budget. Funding UAI via the Arm/Ord PE will ensure that platform and weapon program offices have the support required to implement and update UAI.

This program is in Budget Activity 5 - System Development and Demonstration (SDD) because it supports armament integration, an SDD-type activity.

0604602F references 0605011F Forward Link

0605011F referenced by 0604602F Backward Link

(U) B. Accomplishments/Planned Program ($ in Millions)

	FY 2008	FY 2009	FY 2010
(U) ICD Dev/Updates			5.702
(U) UAI Common Component			0.786
(U) Certification Tool			0.197
(U) Total Cost	0.000	0.000	6.685

This is not a new start; these efforts were performed under PE 0605011F, RDT&E for Aging Aircraft, in FY 2008 and FY 2009.

(U) C. Other Program Funding Summary ($ in Millions)

FY 2008 Actual	FY 2009 Estimate	FY 2010 Estimate	FY 2011 Estimate	FY 2012 Estimate	FY 2013 Estimate	FY 2014 Estimate	FY 2015 Estimate	Cost to Complete	Total Cost

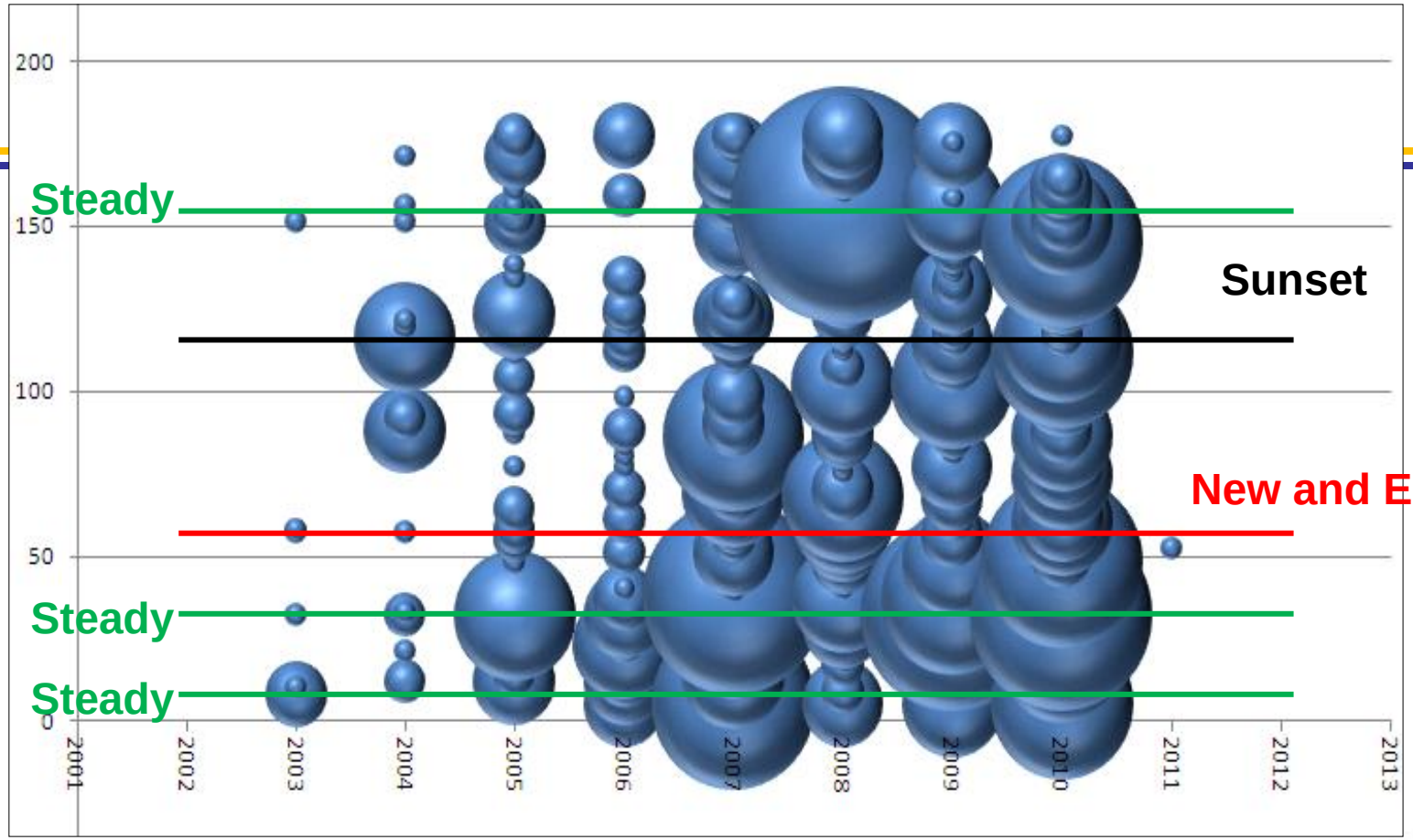
(U) N/A

(U) D. Acquisition Strategy

In December 2004, under the authority of a class Justification and Approval (J&A), the UAI program office awarded individual Cost Plus Fixed Fee (CPFF) contracts to Boeing, Lockheed-Martin, Northrop-Grumman and Raytheon. These four vendors are the Original Equipment Manufacturers (OEMs) for approximately 90% of the Department of Defense' platforms and weapons. Each OEM is responsible for a different piece of the total UAI requirement based on its platform or weapon expertise.

Statistical Significance Tests (Pre-defined Categories)

	Centrality Authority	Radials	Simmelian Ties	Centrality Total Degree	Triad Count	Rank	Value
Growth	0.732	0.481	0.123	0.415	1967.766	2.481	1.104
Die-out	0.665	0.278	0.150	0.478	2646.340	1.423	-1.799
p-value	0.015	0.0015	<0.0001	0.028	0.0002		

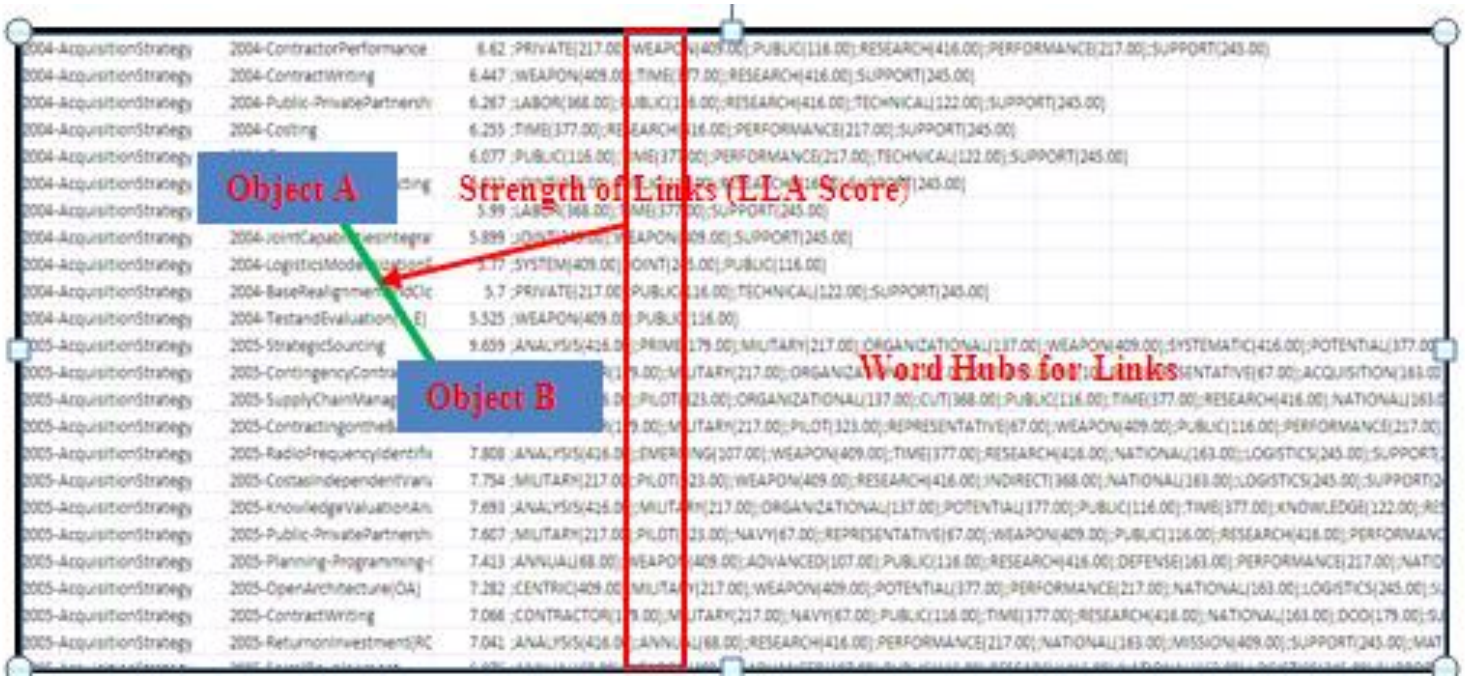


•Steady categories in which the number of reports increased

•New and emerging categories in which there were relatively new.

•Die-down categories in which the number of reports reduced.

Apply LLA to Understand Why Categories Steady, Emerging and Disappearing



•Object: a Year-Category combination

•Link: LLA Score of overlaps of reports for the year and category

Automatic Categories

- Apply LLA to automatically generate themes combined with years as categories
  - 225 of such automatic categories
    - E.g. 2003-COST*COSTS*TOTAL & 2004-SYSTEMS*SYSTEM*PROGRAM
  - We define a value of an automatic category as
    - # of lexical links in the time frame for the theme –

    # of lexical links in the time frame for the same theme
  - Compute the centrality measures for the 225 nodes
    - Links only computed within the same time frame
  - Compute correlation between the centrality measures and "values" of the nodes

e.g. Correlation between "Centrality Authority" and "Value" =0.23 ($p<0.05$ n=225)

Node ID	Centrality Authority/knowledge x knov	Rank	Value
2004-SYSTEMS*SYSTEM*PROGRAM	0.9758	3	94
2004-PERSONNEL*MILITARY*SUPPORT	1	3	90
2004-BUSINESS*INDUSTRY*ARMY	0.7449	3	14
2004-COST*COSTS*TOTAL	0.2685	3	74
2004-CONTRACT*PERFORMANCE*CONTRACTS	0.622	3	29
2003-MODEL*ANALYSIS*APPROACH	0.8503	3	22
2003-PERSONNEL*MILITARY*SUPPORT	0.7443	3	22
2003-SYSTEMS*SYSTEM*PROGRAM	0.525	3	9
2003-PROCESS*PROCESSES*PHASE*PLANNING	1	2	1
2004-SOFTWARE*COMPONENTS*ENGINE*POWER	0.5268	3	25
2004-MANAGEMENT*DECISION*REVIEW	0.3725	3	16

Automatically generated categories

Statistical Significant Correlations Between Centrality and Growth

Pearson Correlation	Centrality Authority (Eigenvalue,PageRank)	Centrality Betweenness	Correlation Expertise	Correlation Resemblance	Centrality Total Degree	Triad Count	Samples	p-value
ARP automatic	0.23	0.24	0.19				225	<0.05
ARP categories			0.15	0.18	-0.12	-0.17	272	<0.05

*Empty cells mean the correlations are not statistically significant

Sort by “Centrality Authority”

arp_automatic_2_cost.xlsx - Microsoft Excel

	A	C	AZ
1	Node ID	Centrality Authority/knowledge x knowledge	Value
2	2007-SYSTEMS*SYSTEM*PROGRAM	1	153
3	2008-SYSTEMS*SYSTEM*PROGRAM	1	-86
4	2009-SYSTEMS*SYSTEM*PROGRAM	1	35
5	2004-PERSONNEL*MILITARY*SUPPORT	1	90
6	2005-SYSTEMS*SYSTEM*PROGRAM	1	25
7	2003-PROCESS*PROCESSES*PHASE*PLANNING	1	1
8	2006-SYSTEMS*SYSTEM*PROGRAM	1	18
9	2004-SYSTEMS*SYSTEM*PROGRAM	0.9758	94
10	2008-PERSONNEL*MILITARY*SUPPORT	0.9258	-27
11	2009-PERSONNEL*MILITARY*SUPPORT	0.9011	48
12	2007-COST*COSTS*TOTAL	0.8795	17
13	2007-PERSONNEL*MILITARY*SUPPORT	0.8629	7
14	2003-MODEL*ANALYSIS*APPROACH	0.8503	22
15	2007-MODEL*ANALYSIS*APPROACH	0.8453	32
16	2003-CONTRACT*PERFORMANCE*CONTRACTS	0.8405	2
17	2009-PROCESS*PROCESSES*PHASE*PLANNING	0.8174	23
18	2007-MANAGEMENT*DECISION*REVIEW	0.8164	1
19	2009-MODEL*ANALYSIS*APPROACH	0.8076	-4
20	2006-PERSONNEL*MILITARY*SUPPORT	0.7956	39
21	2006-COST*COSTS*TOTAL	0.7649	45
22	2009-COST*COSTS*TOTAL	0.7604	31
23	2008-MODEL*ANALYSIS*APPROACH	0.7456	-6
24	2004-BUSINESS*INDUSTRY*ARMY	0.7449	14
25	2003-PERSONNEL*MILITARY*SUPPORT	0.7443	22
26	2006-BASED*PRICE*JOINT	0.7306	9
27	2003-COST*COSTS*TOTAL	0.7256	2
28	2005-PERSONNEL*MILITARY*SUPPORT	0.7173	-15
29	2005-COST*COSTS*TOTAL	0.7126	1
30	2006-MODEL*ANALYSIS*APPROACH	0.7048	69

Sort by “Correlation Expertise”

arp_automatic_2_cost.xlsx - Microsoft Excel

	A	T	AZ
1	Node ID	Correlation Expertise/knowledge x k	Value
2	2004-SYSTEMS*SYSTEM*PROGRAM	0.0329	94
3	2004-BUSINESS*INDUSTRY*ARMY	0.0328	14
4	2004-PERSONNEL*MILITARY*SUPPORT	0.0328	90
5	2004-COST*COSTS*TOTAL	0.0327	74
6	2004-CONTRACT*PERFORMANCE*CONTRACTS	0.0327	29
7	2003-SYSTEMS*SYSTEM*PROGRAM	0.0327	9
8	2003-PERSONNEL*MILITARY*SUPPORT	0.0327	22
9	2003-MODEL*ANALYSIS*APPROACH	0.0327	22
10	2003-PROCESS*PROCESSES*PHASE*PLANNING	0.0327	1
11	2004-SERVED*LCDR	0.0326	11
12	2004-MANAGEMENT*DECISION*REVIEW	0.0326	16
13	2004-SOFTWARE*COMPONENTS*ENGINE*POWER	0.0326	25
14	2003-COST*COSTS*TOTAL	0.0326	2
15	2003-REPORT*REPORTS*ACT	0.0325	1
16	2007-TRAINING*COURSES*INSTRUCTION	0.0325	17
17	2007-SERVED*LCDR	0.0325	15
18	2004-REPORT*REPORTS*ACT	0.0325	15
19	2004-AIR_FORCE*NAVY*AIR	0.0325	12
20	2004-DUE*CONTROL*INVENTORY	0.0325	21
21	2004-BASED*PRICE*JOINT	0.0325	8
22	2003-DATA*INFORMATION*KEY	0.0325	4
23	2004-TIME*SIGNIFICANT*ADDITIONAL*FUNDING	0.0325	53
24	2003-MANAGEMENT*DECISION*REVIEW	0.0325	9
25	2004-MODEL*ANALYSIS*APPROACH	0.0325	55
26	2004-ACQUISITION*DEFENSE*NATIONAL	0.0325	21
27	2003-ACQUISITION*DEFENSE*NATIONAL	0.0325	5
28	2003-PRIOR*COMPETITION*SPECIFIC	0.0325	-4
29	2003-CONTRACT*PERFORMANCE*CONTRACTS	0.0325	2

THEORY

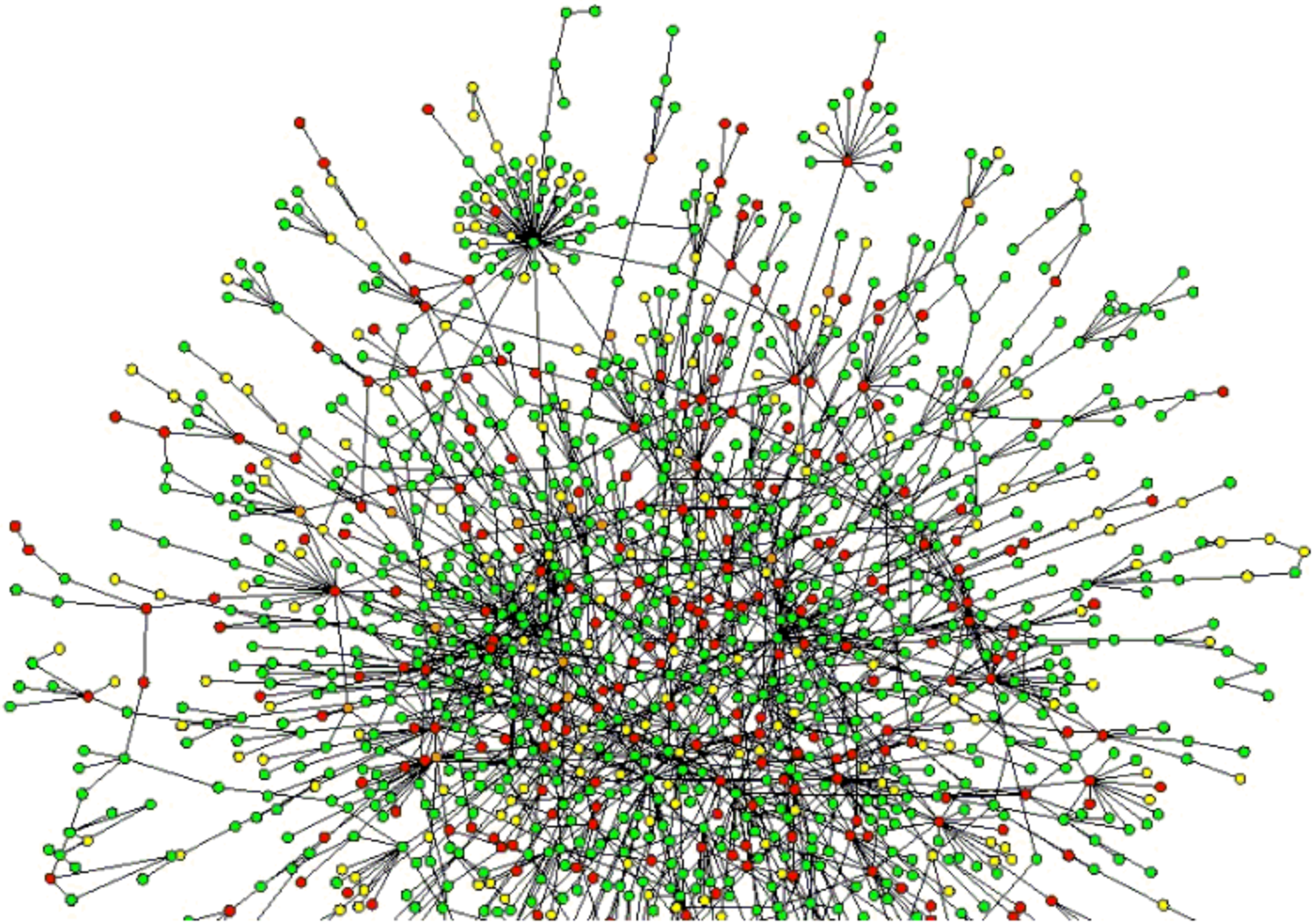
The effect of linguistic constraints on the large scale organization of language

Madhav Krishna[1], Ahmed Hassan[2], Yang Liu[2], Dragomir Radev[2*]
1 Columbia University New York, New York, USA
2 University of Michigan Ann Arbor, Michigan, USA
* E-mail: Corresponding radev@umich.edu

DISE

- Characteristics of a set of important networks and systems of systems
  - WWW , collaboration networks, social networks, US power grid, metabolic networks, semantic networks,
  - Share the same characteristics
    - Power-law, scale-free: relatively small number of well-connected nodes serve as hubs Pareto principle, 80/20 rule
    - Small-world phenomenon (random two nodes ,e.g. two person in US, only separated by six degrees away)
    - Self-organizing
    - Self similar (fractals)
    - Preferential attachment

The E.coli metabolic network is scalefree (PZM Pareto-Zipf-Mandelbrot type, parabolic fractal) and has small-world properties



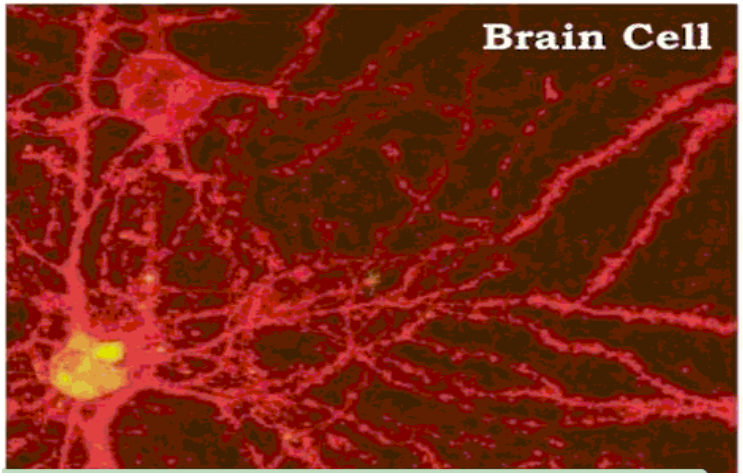
http://www.bordalierinstitute.com/targetted.html Connect to fractals?

bordalier institute

contact: winiwarter@bordalierinstitute.com

Research scope : complex systems / neural networks & evolution

"*I think the next century (21st) will be the century of complexity.*" Stephen Hawking



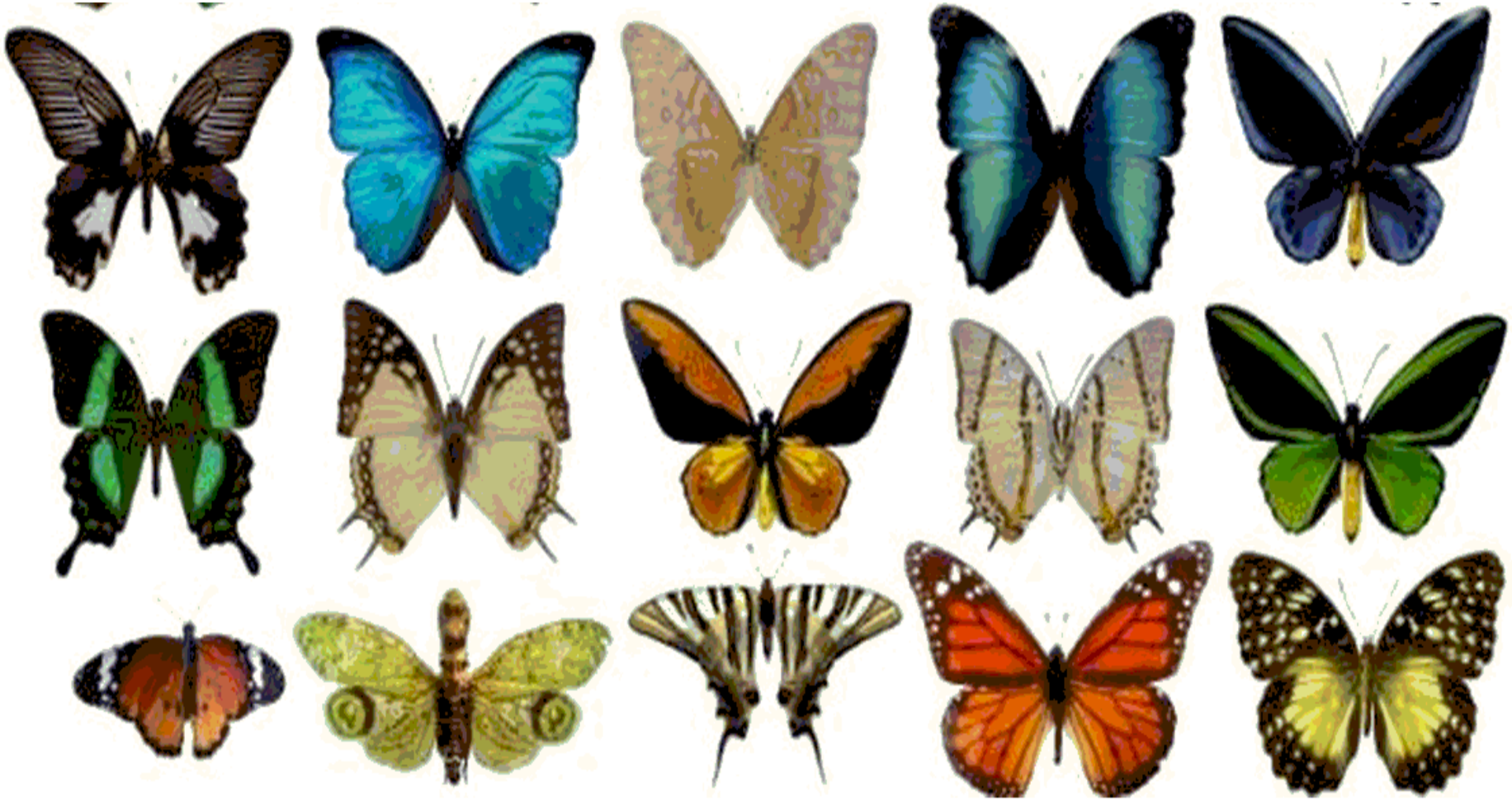
en.wikipedia.org/wiki/File:Water_Crystals_on_Mercury_20Feb2010_CU1.jpg

File Talk

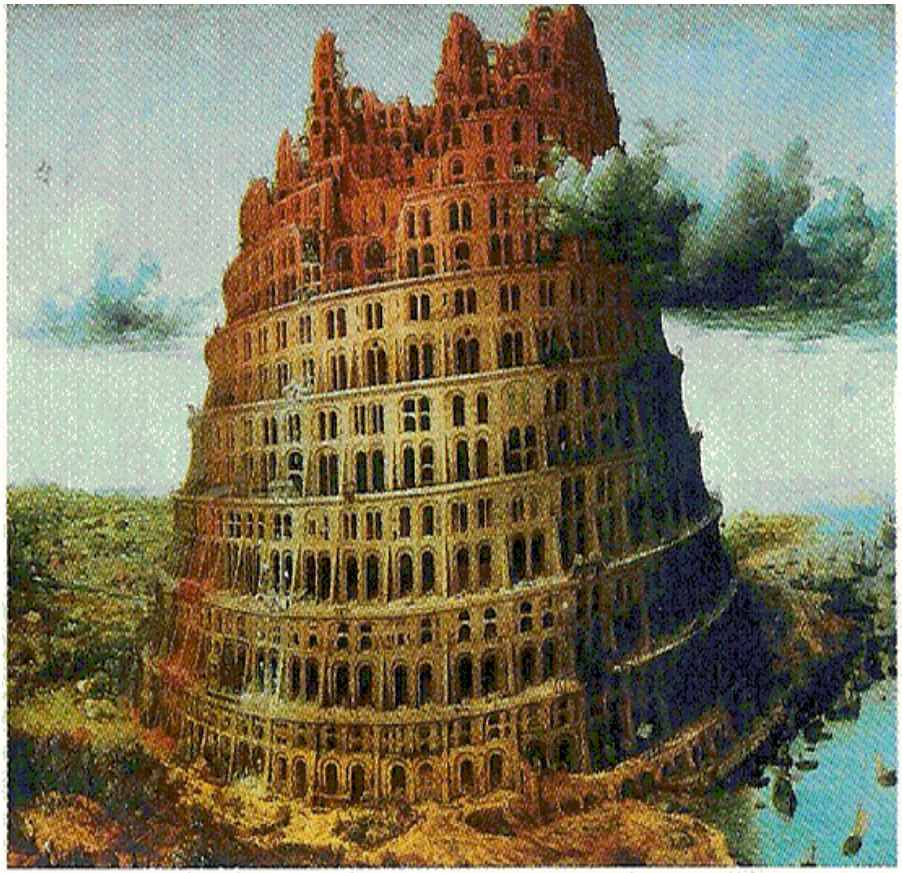
File:Water Crystals on Mercury 20Feb2010 CU1.jpg

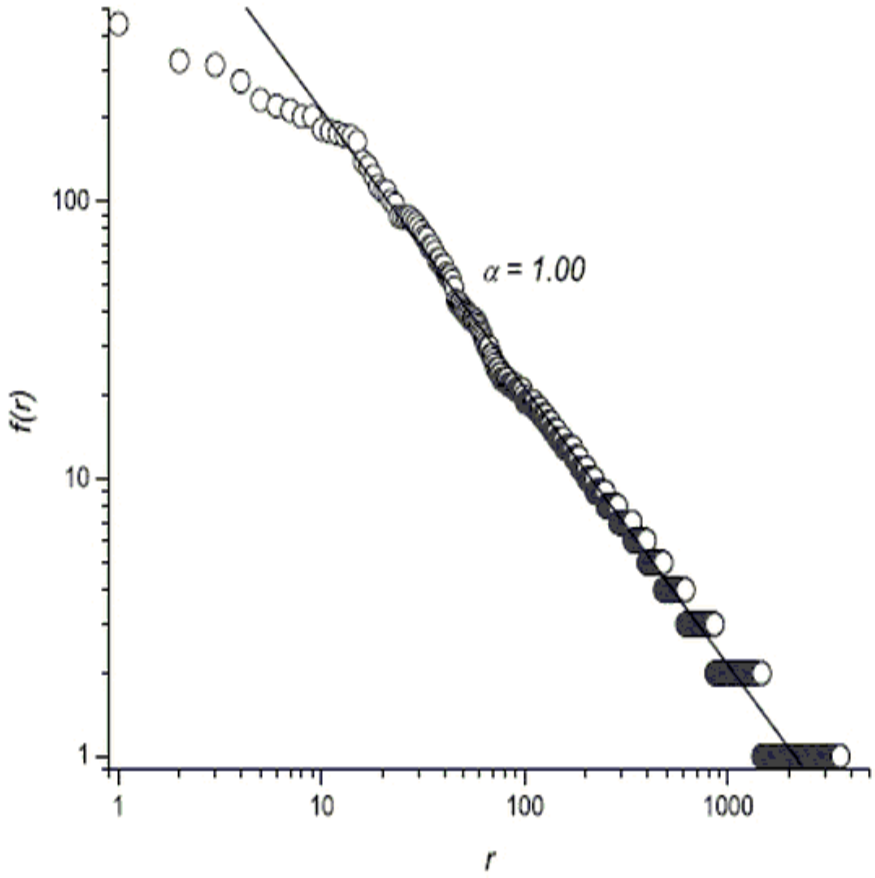
From Wikipedia, the free encyclopedia

File File history File usage



A zebra's stripes, a seashell's spirals, a butterfly's wings: these are all examples of patterns in nature. The formation of patterns is a puzzle for mathematicians and biologists alike. How does the delicate design of a butterfly's wings come from a single fertilized egg? How does pattern emerge out of no pattern?
http://www.sciencedaily.com/releases/2008/06/080619111748.htm

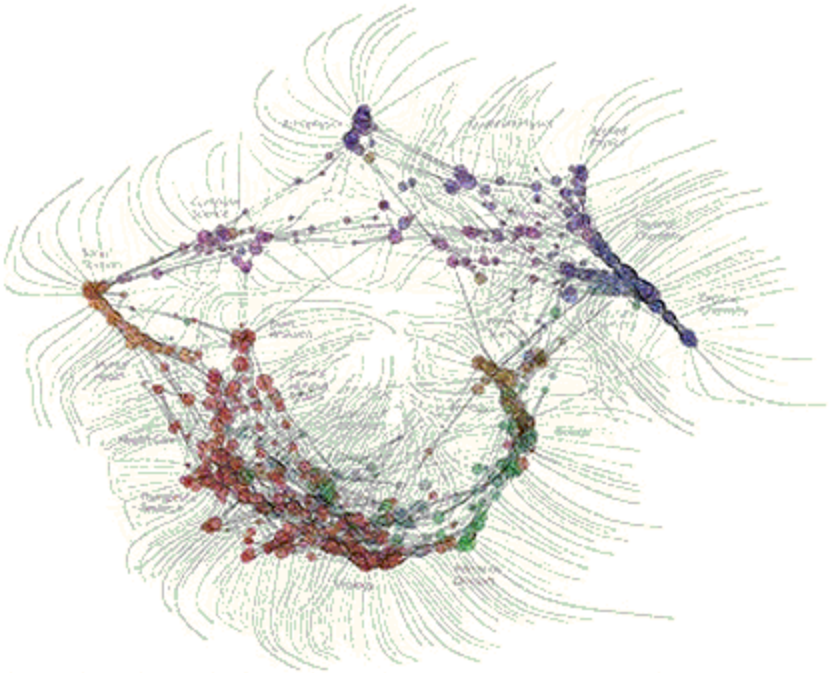




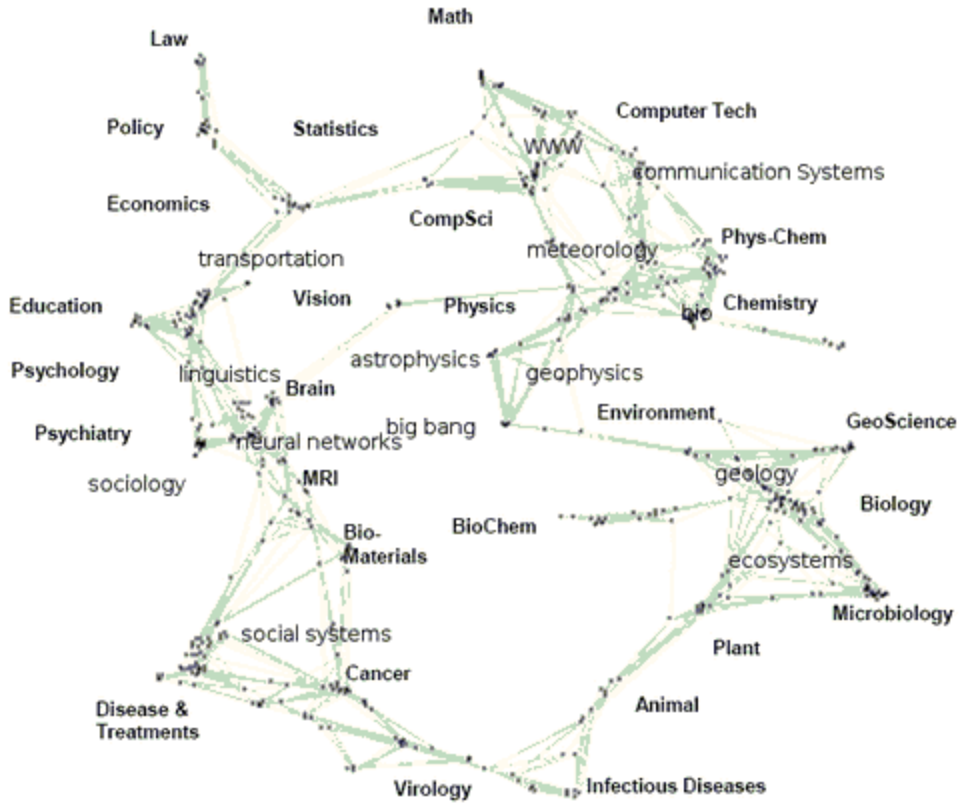
word frequency distributions (Zipf's law)

PZM (Pareto-Zipf-Mandelbrot, parabolic fractal) distributions are observed for the word frequencies of all texts of all languages, all times, any age of author, even the bubbling of babies show the same Pareto-Zipf-Mandelbrot distribution with a slope of 1.00 ...

- scientometrics,
  research and publications :



Research & Node Layout: Kevin Boyack and Dick Klavans (mapofscience.com);
Data: Thompson ISI; Graphics & Typography: W. Bradford Paley (didi.com/brad);
Commissioned Katy Börner (scimaps.org)



topic map of science with links (click on the subject) to PZM Pareto-Zipf-Mandelbrot (parabolic fractal) distributions.

Self-organizing

- A system of elements spontaneously forming of well organized structures[de Boer, 1998]
  - Elements are distributed i.e., no single element coordinates the activity
  - Patterns, or behaviors, from random initial conditions.
  - Self limiting, limits its own growth by its actions
  - Universal mechanism for social animals and simple mathematical structures, expected in human society. e.g. the wireless communications industry.
  - Tell-tale signs of self-organization are
    - statistical properties shared with self-organizing physical systems (i.e. Zipf's law, power-law, Pareto principle).
- Emerge from bottom-up interactions, and appear to be limitless in size. Top-down hierarchical networks, which are not self-organizing.
- In economics,
  - Market economy is sometimes said to be [Krugman,1996].
  - Friedrich Hayek coined the term catallaxy as to exchange, to admit in the community and to change from enemy into friend, which is an alternative expression for the word economy, now a new dimension in software design and network architecture [Eymann, Padovan & Schoder, 2000], to describe a "self-organizing system of voluntary co-operation."
  - Central planning is not and less efficient.

Growth Theories Using Centrality

- Degree-based centrality,
- In-degree, out-degree and total degree,
- Google's PageRank algorithm among others such as
    - hub and authority centralities belongs to this group.
- A betweeness centrality describes whether and how frequently a node is part of the shortest paths between pairs of nodes in the network.
- A closeness centrality is defined in terms of the lengths of the shortest paths from a node to the rest of the nodes in the networks.
- ***Structure Holes***[Burt, 2005]
    - Structural holes refer to the absence of ties between two parts of a network.
    - Finding and exploiting a structural hole can give an entrepreneur a competitive advantage. Ronald Burt, 1995, 2005], and is sometimes referred to as an alternate conception of social capital
    - Actors with a lot of structural holes (i.e. nonredundant ties) in their network are supposed to hold informational and control advantages that allow them to capitalize from their social networks in ways that others cannot. These people occupy a brokering position. The standard argument is that a network with many structural holes leads to better financial outcomes, greater returns to investment, etc.
    - But it's possible that the standard theory of structural holes is based on an individualistic, Western view of human behavior. That is, it assumes that people adhere to the individualistic principles of Western culture. What happens to people with networks rich in structural holes that live/work in environments that adhere to other principles, such as those of a collectivistic culture?
    - http://orgtheory.wordpress.com/2007/06/19/structural-holes-in-context/

Preferential Attachment (PA) [Barabási & Albert, 1999]

- **The most popular explanation**
  - a new node is connected to a pre-existing one with a probability proportional to the number of links (degree) of the target node
  - any of a class of processes in which some quantity, e.g. wealth or credit, is distributed among a number of individuals or objects according to how much they already have, so that those who are already wealthy receive more than those who are not.
  - ‘rich get richer’ ,
  - "Yule process",
  - "cumulative advantage",
  - the "Matthew effect".
  - the first application of the process was to grow a random network to a scale-free network[Price, 1976]. Price also promoted preferential attachment as a possible explanation for power laws in many other phenomena
  - Lotka's law of scientific productivity
  - Bradford's law of journal use,
  - Gibrat's law of business or firm growth
  - Zipf's law of city sizes.
- **Successful in predicting the graph structure of the web among others**
- **Problems with PA**
  - As time evolves, new nodes join the network by adding links with a probability proportional to the degree of existing nodes.
  - Higher degree of a node reflects higher relevance or popularity.
  - Earlier nodes tend to have significantly higher degrees than later ones, making it hard for a node which enters late to compete with the already established hubs of the network[Borgs, Chayes, Daskalakis & Roch, 2007].

Pareto Optimal

- Pareto efficient
  - Given an initial allocation of goods among a set of individuals, a change to a different allocation that makes at least one individual better off without making any other individual worse off is called a *Pareto improvement*.
  - An allocation is defined as "Pareto efficient" or "Pareto optimal" when no further Pareto improvements can be made.
- A system that is *not* Pareto efficient
  - implies that a certain change in allocation of goods (for example) may result in some individuals being made better off with no individual being made worse off, and therefore can be made more Pareto efficient through a Pareto improvement.
  - Here *better off* is often interpreted as put in a preferred position, for example, more central or higher degree
- Implications
  - Game theory: <the problem of a coordination failure>
    - The existence of externalities lead to coordination failure and results in Pareto-inferior outcomes.
  - Computer science: <the price of anarchy>
    - Selfish behavior may not achieve full efficiency at the collective level.

http://arxiv.org/pdf/nlin/0502003.pdf

SWARMING: NETWORK ENABLED C4ISR © 2003 BY MARK FLEISCHER. ALL RIGHTS RESERVED. 1

Foundations of Swarm Intelligence:
From Principles to Practice

Flocking Behaviors

What is the mechanism behind flocking behavior?

oach is based on what is often referred
:e (SI). The term SI has come to repre-
ossible to control and manage complex
entities even though the interactions
e entities being controlled is, in some
otion therefore lends itself to forms of
may be much more efficient, scalable
complex systems.
ures of SI are based on observations
colonies and beehives, for example,
roperty that large numbers of them
affairs in a very organized way with
behavior that enhances their collective
nd paradoxically, these insects seem to
es of interaction. This phenomenon is
ddressed in other domains of inquiry
involving *complexity* such as cellular automata and the study

preserved under changes in syste ates. This provides a

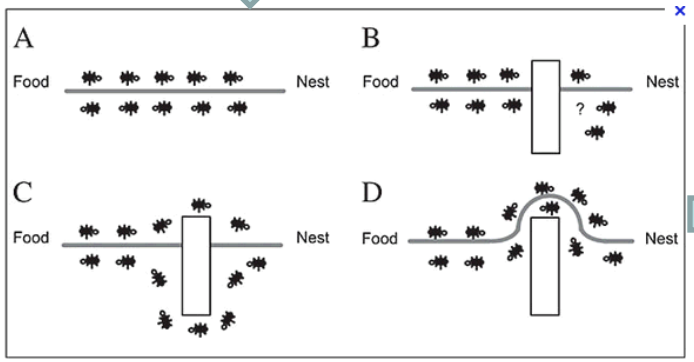


Figure 2. A. Ants in a pheromone trail between nest and food; **B.** an obstacle interrupts the trail; **C.** ants find two paths to go around the obstacle; **D.** a new pheromone trail is formed along the shorter path.

- Self-organized to collective better;
- Local, simple communications but achieves Pareto optimal

(http://www.funpecrp.com.br/gmr/year2005/vol3-4/wob09_full_text.htm)

- Use for design armed forces, wireless communications, cellular automata, peer-to-peer networks where one wants to have strong collective intelligence for the whole network/system

shorter paths have a stronger increment in pheromone

Collaborative Learning Agents

At any given time, we are able to rank the knowledge themes based on its predicted future importance, and distribute themes among stakeholders and social actors.

•Measure the fitness of the whole system. On a theoretic level, we will

•Hidden Markov Models (HMM) for global optimization with a local learning:

Observations O(t): Characteristics about a single agent/actor/ that is observable, e.g. measures of single stakeholder's awareness of information using lexical links;

Hidden state *j*, *j=1,…J,* Hidden information that is interesting but difficult to observe directly from data, e.g. stakeholders and regulators can possess different types of competitiveness, reward.

We will also model the predictive relation between lexical links *O(t)* and *hidden* states as a probability density function *bj(O(t))* = *b(a(t)=aj|O(t))*. The overall fitness *R(t,aj)* means the total fitness of a complex system up to time *t*. The overall fitness function can be computed recursively.

bj(O(t))

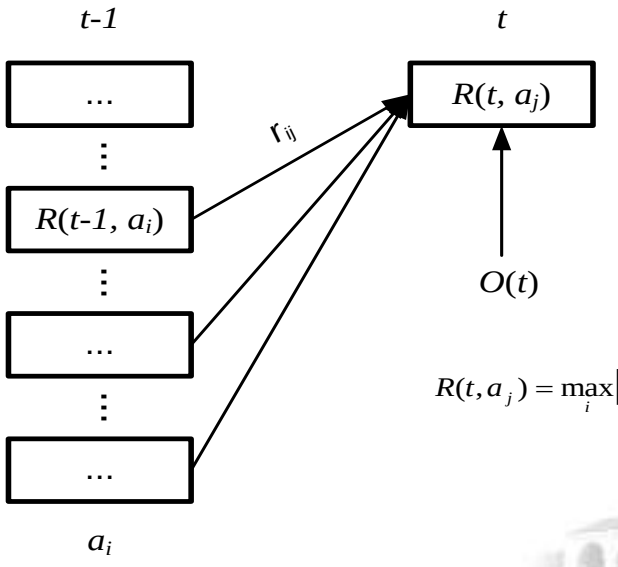
-Measure of reward of a single agent action with the local knowledge of

-e.g. **self-awareness of an individual actor on how different, diversified, anomalous the agent is from others.**

-*R(t,aj)* a global fitness

-Multi-agent systems

Recursion to Compute the Overall Fitness of a System *R(t, aj)*

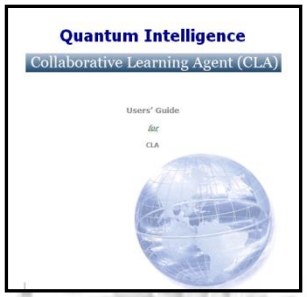


$$R(t,a_j) = \max_i \left[R(t-1,a_i) + r_{ij}\right] + b_j\left[O(t)\right]$$

BACKUP

Table of Contents

Table of Contents



Lexical Link Analysis

- Lexical Link Analysis (LLA) is a form of text analysis
  - A text is represented as a network of lexical terms (e.g. word pairs, bigram) if they are in a community of a word network.
  - Word pairs are further grouped into concepts and themes using large-scale social network community detection algorithms
  - Consequently the importance, impact and evolution of these concepts and themes can be revealed, as well as the crucial relationships among pre-defined categories or automated discovered clusters.
- In a nutshell, LLA is a statistical co-occurrence, bi-gram TAN method for text analysis.
  - *Singlish* (Singapore English mixed English and Chinese)
  - Biological systems within their own symbols for representations.
  - We want to emphasize the connection of LLA's connection to the theories and practices of complex systems and systems of systems, where anticipated benefits of such analysis and presentation are manifested into the concept of System Self-awareness.
- Core focus: Use LLA to automatically discover the concepts and themes in large-scale texts and represent them as dynamic evolving networks over time
  - As a new way to predict the emergence of new information.
  - Discuss the relationship of LLA to complex system theories and network centrality measures.
  - Use cases examine the content of diversified unstructured data, identify new information that might have large impacts and growth potentials in the future.

How LLA Computed

- Read each set of documents.
- Select feature-like word pairs.
- Apply a social network community finding algorithm (e.g. Newman grouping method; Girvan et al. 2001) to group the word pairs into themes. A theme includes a collection of lexical word pairs connected each other.
- Compute a “weight” for a theme for the information of a time period, that is, how many word pairs belong to a theme for that time period and for all the time periods.
- Sort theme weights by time, and study the distributions of the themes by time.
- General questions that LLA usually answers are as follows:
  - Discover themes and topics in the unstructured documents and sort the importance of the themes
  - Discover social and semantic networks of organizations who were involved, compare the two networks to obtain insights to answer the following questions:
  - What were the organizations involved in the *important* themes
  - How do semantic networks suggest more potential collaboration when compared to social networks?

Text Analysis/Mining Tasks

- Named Entity Extraction (NEE)
  - People, place, date, money, etc.
- Text Summary
- Text Categorization
- Text Clustering
- Concept Extraction
- Topic/Theme Extraction
- Text Dynamics: Emergence of New Concepts/Themes Over Time
- Sorting documents, keywords and themes
  - Search

Cognitive Expertise

Informal Name

None

Citation

Carley, Kathleen M. 2002. "Summary of Key Network Measures for Characterizing Organizational Architectures." Unpublished Document: CMU 2002

Description

Measures the degree to which each pair of agents has complementary knowledge, expressed as a percentage of the knowledge of the first agent.

Example : If one person in an organization knows how do perform X but can't do Y. Whereas another individual can do Y but not X, such individuals would rank highly in this measure.

Input : Agent (source) by Knowledge (knowledge) matrix with DataType=binary.

Output : $\Re \in [0,1]$

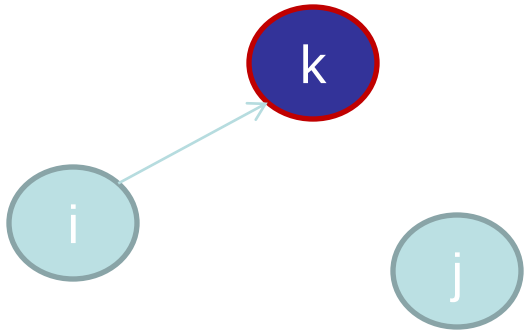
Node Level with Type=agent and DataType=real.

Dyad Level with Type=agent, Target=agent, and DataType=real.

For each pair of agents (i,j) compute the number of knowledge bits that j knows that i does not know. Then normalize this sum by the total number of knowledge bits that agent i does not know.

$$CE_{i,j} = \frac{\sum_{k=1}^{|K|} (\sim AK_{i,k} * AK_{j,k})}{(|K| - \sum_{k=1}^{|K|} AK_{i,k})}$$

$$CE_{i,i} = 0$$



NOTE : The CD output matrix is NOT-symmetric.